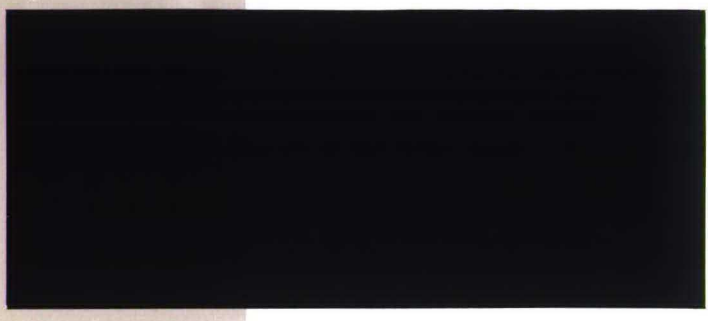
entER for mic Research

Discussion paper



TILBURG UNIVERSITY

No. 9107

REFINEMENTS OF NASH EQUILIBRIUM

by Eric van Damme

February 1991

ISSN 0924-7815

Refinements of Nash Equilibrium*

Eric van Damme†

1 Introduction

Noncooperative game theory studies the question of what constitutes rational behavior in situations of strategic interaction in which players cannot communicate nor sign binding agreements. The traditional answer to this question centers around the notion of Nash equilibrium. Such an equilibrium is a vector of strategies, one for each player in the game, with the property that no single player can increase his payoff by changing to a different strategy as long as the opponents do not change their strategies. The Nash equilibrium concept is motivated by the idea that a theory of rational decision making should not be a self-destroying prophecy that creates an incentive to deviate for those who believe in it. To quote from Luce and Raiffa (1957, p. 173)

> "if our non-cooperative theory is to lead to an n-tuple of strategy choices and if it is to have the property that knowledge of the theory does not lead one to make a choice different from that dictated by the theory, then the strategies isolated by the theory must be equilibrium points."

In other words, for a (commonly known) norm of behavior to be self-enforcing it is *necessary* that the norm (agreement) constitutes a Nash equilibrium.

---

*Paper presented at the 6th World Congress of the Econometric Society, Barcelona, 22-28 August, 1990. The author thanks Helmut Bester, Larry Samuelson and Jonathan Thomas for comments on an earlier version.

†CentER for Economic Research, Tilburg University, the Netherlands

The increased use of noncooperative game theory in economics in the last decades has led to an increased awareness of the fact that not every Nash equilibrium can be considered as a self-enforcing norm of behavior. Very roughly, the Nash concept is unsatisfactory since it may prescribe irrational behavior in contingencies that arise when somebody has deviated from the norm. In applications, one typically finds many equilibria and intuitive, context depending arguments have been used to exclude the 'unreasonable' ones. At the same time game theorists have tried to formalize and unify the intuitions conveyed by applications and examples by means of general refined equilibrium notions. The aim of this paper is to describe, and comment on the most important concepts that have been put forward as being necessary for self-enforcingness. Although the literature offers a wide variety of different refinements, it will be seen that all of them are based on a small number of basic ideas. (These main ideas are also described in Kohlberg (1989) from which I borrowed the term "norm of behavior".)

Ever since Luce and Raiffa (1957) the intuitive justification of equilibria and the relevance of equilibria to the analysis of a game have been questioned. It has been realised that it is not evident that Nash equilibrium is a necessary consequence of strategic reasoning by rational players, that it is not clear how players would arrive at an equilibrium or how they would select one from the set of equilibria. I do not wish to enter a discussion on these topics here, rather I refer to Aumann (1987a, 1988), Bernheim (1986), Binmore (1990), Brandenburger and Dekel (1987) and Tan and Werlang (1988) for extensive discussions on the epistemic foundations of equilibria, i.e. on what the players must know about each other's strategies and each other's rationality for equilibria to make sense. In this author's opinion some of the confusion surrounding the Nash concept can be traced to the fact that the mathematical formalism of noncooperative game theory allows multiple interpretations and to the fact that the different aspects of noncooperative analysis are not clearly separated.

Noncooperative game theoretic analysis has several aspects:

(i) (The equilibrium definition problem.) Which agreements are self-enforcing?

(ii) (The equilibrium attainment problem.) How, or under which conditions will the players reach an agreement?

(iii) (The equilibrium selection problem.) Which agreement is likely to be concluded?

Except for the last section I deal exclusively with the first topic. I do not discuss how self-enforcing norms come to be established nor how the selection among these takes place. The motivation for studying the first question independently is that knowing its answer seems a prerequisite for being able to answer the other questions. (For example, one might hope that in games with a unique self-enforcing equilibrium players will always coordinate on that equilibrium.) I restrict attention to refinements of Nash equilibrium that try to capture further necessary conditions for self-enforcing norms of behavior. Hence, I investigate which conditions Nash equilibria should satisfy such that rational players would have no incentive to deviate from them. Using the terminology of Binmore (1987) I, therefore, remain in the eductive context.

Nash equilibria also admit other interpretations than as self-enforcing norms and in other (non-eductive) contexts different considerations, leading to alternative refinements, may be appropriate. For example, in biology an equilibrium is seen as the outcome of a dynamic process of natural selection rather than as the consequence of reasoning by the players. The basic equilibrium concept in that branch of game theory, viz. the notion of evolutionarily stable strategies or ESS (Maynard Smith and Price (1973), Maynard Smith (1982)) may formally be viewed as a refinement of Nash equilibrium but it is not further discussed here since it is motivated completely differently. (Although, mathematically it is related to several concepts discussed below, see Van Damme (1987, Chapter 9).) Similarly I will not deal with the interpretation of Nash equilibria as stable states of learning processes in a context in which the same game is played repeatedly, but each time with different active players who can use observations from the past to guide their behavior. (On learning models, see, for example, Canning (1989, 1990), Fudenberg and Kreps (1988), Kalai and Lehrer (1990) and Milgrom and Roberts (1989, 1990)).) Of course, this does not imply that I consider such contexts to be unimportant, they simply fall outside the scope of this paper. Perhaps in economic situations learning and

evolution are even more important than reasoning. Finally, I rule out any correlation between players' actions that is not explicitly allowed by the rules, hence, I do not consider correlated equilibria (Aumann (1974), Forges (1986), Myerson (1986)).

Space limitations do not allow an extensive discussion on the applications of the various refinements. Yet, the proof of the pudding is in the eating, it is the applications and the insights derived from them that lend the refinements their validity. As Aumann (1987b) writes

> "My main thesis is that a solution concept should be judged more by what it does than by what it is; more by its success in establishing relationships and providing insights into the workings of the social processes to which it is applied than by considerations of *a priori* plausibility based on its definition alone."

The remainder of the paper is organised as follows. In Section 2 I discuss the principle of backward induction, i.e. the idea that an equilibrium strategy should also make sense in contingencies that do not arise during the actual play. Special emphasis is on the concepts of subgame perfect and sequential equilibria, on the definition of consistency of beliefs and on the assumption of persistent rationality. Section 3 deals with "trembling hand perfect" equilibria as well as the related notions of properness and persistency. All three concepts require that the equilibrium still makes sense if with a small probability each player makes a mistake. This section also briefly investigates what kind of refinements result if it is required that an equilibrium be robust against slight perturbations in the payoffs or in the structure of the game. Issues related to the Kohlberg/Mertens concept of stable equilibria are discussed in section 4. Stability is a set-valued solution concept and it will be shown that set-valuedness is a natural consequence of several desirable properties. The topic of Section 5 is forward induction, i.e. the idea that a player's past behavior may signal either this player's private information or how the player intends to play in the future. For the special class of signalling games several

intuitive refinement criteria are reviewed that are all related to Kohlberg/Mertens stability. In Section 6 we move from equilibrium refinement to equilibrium selection and briefly discuss a model (originally due to Carlsson and Van Damme) in which slight payoff uncertainty forces players to coordinate on a specific 'focal' equilibrium in each 2×2 bimatrix game.

This introduction is concluded by specifying the notational conventions that will be used for extensive form games. Attention will be confined to finite games with perfect recall and for the definition of such a game Γ the reader is referred to Selten (1975) or to Kreps and Wilson (1982a). X denotes the set of decision points in Γ, Z is the set of endpoints and $u_i(z)$ is player i's payoff when z is reached. We depict the endpoints by row-vectors, the first component of which is the payoff to player 1, etc. The origin of the game tree is depicted by an open circle. H_i denotes the set of information sets of player i (with typical element h). We depict an information set by a dashed line that connects the points in the set. A behavior strategy s_i of player i assigns a local strategy s_{ih} (i.e. a probability distribution on the set of choices at h) to each $h \in H_i$. If s is a (behavior) strategy vector, $s = (s_1, \ldots, s_n)$, then p^s, the outcome of s, is the probability distribution that s induces on the set of endpoints of Γ. If A is a set of nodes, we also write $p^s(A)$ for the probability that A is reached when s is played. Player i's expected payoff resulting from s is denoted by $u_i(s)$, hence, $u_i(s) = \sum_z p^s(z)u_i(z)$. For a decision point $x \in X$, denote by p^s_x the probability distribution that s would induce on Z if the game were started at x, and write $u_{ix}(s) = \sum_z p^s_x(z)u_i(z)$. If μ specifies a probability distribution on the decision points in the information set $h \in H_i$, then we write $u^\mu_{ih}(s) = \sum_{x \in h} \mu(x)u_{ix}(s)$. If s is a strategy vector and s'_i is a strategy of player i, then $s \backslash s'_i$ denotes the strategy vector $(s_1, \ldots, s_{i-1}, s'_i, s_{i+1}, \ldots, s_n)$, We use S_i to denote the set of all strategies of player i and S is the set of strategy vectors.

2 Backward Induction

A strategy vector s is a Nash equilibrium (Nash (1950a)) of an extensive form game Γ if

$$u_i(s) \geq u_i(s\backslash s_i') \quad \text{for all } i \text{ and all } s_i' \in S_i. \tag{2.1}$$

If we interpret a strategy vector s as a (fully specified) norm of behavior then (2.1) is a necessary condition for a commonly known norm to be self-enforcing, i.e. for the norm to be such that no player has an incentive to deviate from it. In this interpretation, s_{ih} (the local strategy of player i at h) may be viewed both as player i's intended action at h as well as the common prediction of all the opponents of what i will do at h. (For further comments on the interpretation of strategies, see Rubinstein (1988).) Hence, Nash equilibrium requires common and correct conjectures. It is important to note that, for a Nash equilibrium, it is necessary that different players conjecture the same response even at information sets that are not reached when s is played. (Cf. the discussion on the game of Figure 11 in Section 5.2.) In extensive form games, taking strategy vectors as the primitive concept in particular implies that a player's predictions do not change during the game: Player j's conjecture about the action chosen at h is s_{ih} both at the beginning of the game as well as at any information set $k \in H_j$, even if it is the case that k cannot be reached when s_i is played. Hence, taking strategy vectors as the primitive concept implies an assumption of "no strategy updating", i.e. that at each point in time each player believes that in the 'future' all players will behave according to the norm even though he may have seen that players did not observe the norm in the past. We make these remarks to show that some criticisms that have been leveled against subgame perfect equilibria are actually criticisms against using strategy vectors as the primitive concept of a theory.

2.1 Subgame perfect equilibria

Selten (1965) provided an example similar to the game from Figure 1a to point out that not every Nash equilibrium can be considered a self-enforcing norm of behavior: (D, d) is a Nash equilibrium (player 1 optimises by choosing D if player 2 chooses d and, if player 1 indeed chooses D then player 2's choice is irrelevant since he doesn't have to move).

However, since player 2 cannot commit himself to his choice of d (the game is assumed to be noncooperative), he will deviate to a if he is actually called to play. Even if there is a prior agreement to play (D, d), player 1 anticipates that player 2 will deviate and he deviates as well, thereby increasing his payoff: The agreement is not self-enforcing.

[Insert Figure 1 here]

Nash equilibrium requires that each player's strategy be optimal from the ex ante point of view. Ex ante optimality implies that the strategy is also optimal in each contingency that arises with positive probability but, as the example shows, a Nash equilibrium strategy need not be a best reply at an information set that initially is assigned probability zero. A natural suggestion is to impose ex post optimality as a necessary requirement for self-enforcingness. For games with perfect information (i.e. games in which all information sets are singletons) this requirement of sequential rationality is mathematically meaningful and may be formalized as in (2.2).

$$u_{ih}(s) \geq u_{ih}(s \backslash s_i') \text{ for all } i, \text{ all } s_i' \in S_i, \text{ all } h \in H_i, \tag{2.2}$$

hence, at each information set h player i's equilibrium strategy maximizes the player's expected payoff conditional on having reached h as long as the opponents play their equilibrium strategies in the future. Clearly, equilibria satisfying (2.2) can be found by rolling back the game tree in a dynamic programming fashion. Selten (1965) noted that the argument leading to (2.2) can be extended to a wider class of games. Define a *subgame* as a part of the tree of an extensive form game that constitutes a game in itself. Selten argued that a self-enforcing norm should induce a self-enforcing norm in each subgame since otherwise some player might find it advantageous to deviate from the norm and thereby reach a subgame with an outcome that benefits him. Selten defined

a *subgame perfect equilibrium* as a Nash equilibrium that induces a Nash equilibrium in every subgame.

In condition (2.2) it is assumed that each player at each point in time believes that in the future all players will try to maximize their payoffs. A player is required to have such beliefs even in situations in which he has already seen that some players did not maximize in the past: The information set h may be reached only if a deviation from s has occurred. This assumption of persistent rationality has been extensively criticized in the literature (see, for example, Basu (1988, 1990), Binmore (1987), Reny (1988a, b) and Rosenthal (1981)). The critique may be illustrated by means of the game of Fig. 1.b. As long as $x > 1$, the unique strategy vector satisfying (2.2) is (A, D_2a). However, if $x = 4$, then A_2 is strictly dominated so that player 1 only has to move after player 2 has taken an irrational action. In such a situation it is not compelling to force player 1 to believe that player 2 will certainly behave rationally and play a at his second move. There seems no convincing argument why player 1 could not believe that player 2 will choose d, and in the latter case he would prefer D. Reny (1988a) proposes to weaken (2.2) by demanding optimising behavior of player i only at information sets h that are not excluded by player i's own strategy, i.e. that do not contradict the rationality of player i. Reny's concept of 'weak sequential equilibrium' does not put any restrictions on the conjectures about player i's behavior at information sets $h \in H_i$ that can be reached only when player i deviates from s_i. In the game $\Gamma_2(x)$ with $x > 1$ there are multiple weakly sequential equilibria but they all lead to the outcome (x, x). If, however, the game would be modified such that the payoff after A_2Aa would be (4,4) rather than (3,1), then (D, D_2d) would be a weak sequential equilibrium of $\Gamma_2(1.5)$ and this produces an outcome that differs from the subgame perfect equilibrium outcome. (In the modified game Reny's concept allows player 1 to believe that player 2 will choose d after a defection to A_2.)

In Kohlberg and Mertens (1986) it is also proposed to weaken requirement (2.2). These authors take the position that requiring a theory of rationality to specify a unique

choice in every contingency is unduly restrictive and they propose (certain) sets of strategy vectors (rather than single strategy vectors) as the primitive concept of a theory. Hence, according to Kohlberg and Mertens, a self-enforcing norm need not completely pin down the players' behavior and beliefs in those contingencies that will not be reached when the norm is obeyed; we may be satisfied if we can identify the self-enforcing outcomes, i.e. the outcomes that result when everybody obeys the norm. For example, in $\Gamma_2(4)$ the norm that says "player 2 should play D_2" (without specifying what player 1 should do) is self-enforcing in the more liberal sense. In $\Gamma_2(2)$, Kohlberg and Mertens also identify player 2 choosing D_2 as the self-enforcing outcome but now player 1's behavior cannot be completely arbitrary: A self-enforcing norm specifies that player 1 should choose D with a probability of at most $\frac{1}{2}$ since otherwise player 2 will violate the norm. We will return to the Kohlberg/Mertens stability concept in Section 5. In that section it will be seen that several desirable properties that we might want self-enforcing norms to possess can only be satisfied by norms that allow some freedom of choice in some circumstances.

The example from Figure 1.b makes clear that the assumptions that players are perfectly rational and that the game is exactly as specified imply that counterfactuals arise naturally in game theory. As Selten and Leopold (1982) write

> "In order to see whether a certain course of action is optimal it is often necessary to look at situations which would arise if something non-optimal were done. Since in fact a rational decision maker will not take a non-optimal choice, the examination of the consequence of such choices will necessarily invoke counterfactuals."

In the game $\Gamma_2(4)$, to determine his optimal choice, player 1 has to evaluate the counterfactual "if player 2 would choose A_2, my best response would be A". Philosophers (Lewis (1973) and Stalnaker (1969)) have suggested evaluating such a counterfactual by investigating whether in a world (or model) that is most similar to the one under consideration and in which player 2 chooses A_2 it is indeed true that the best response is A.

Selten and Leopold (1982) suggest a parameter theory of counterfactuals, a slight variation of this idea. To implement this idea, game theorists have suggested to formalize the similarity relation by means of perturbed games: the original game is embedded into a larger perturbed game (in which all information sets are reached) and is approximated by letting the perturbations vanish. Two possible perturbations readily suggest themselves, one may either give up the assumption that the players are perfectly rational (this is the approach taken in Selten's perfectness concept, see subsection 3.1) or one may give up the assumption that the game model fully describes the situation. Some consequences of the latter approach will be investigated in subsection 3.2. Not surprisingly, it will be seen that different approaches may yield different outcomes.

Before turning to perturbations, however, we first discuss the concept of sequential equilibria.

2.2 Sequential Equilibria

The ex post optimality requirement (2.2) cannot be applied at non-singleton information sets since there the conditional expected payoff need not be well-defined. As a consequence, the requirement of subgame perfection does not suffice to rule out all non-self enforcing equilibria. For example, change the game from Figure 1.a such that player 1 chooses between D, A and A' with player 2 moving after A and A' but without knowing whether A or A' was chosen and with the payoffs after A' being the same as those after A. Then (D, d) is a subgame perfect equilibrium of the modified game (since the latter admits no subgames), but it clearly is not self-enforcing.

Kreps and Wilson (1982a) suggest extending the applicability of (2.2) by explicitly specifying beliefs (i.e. conditional probabilities) at every information set so that posterior expected payoffs can always be computed and they propose to make these beliefs a formal part of the definition of an equilibrium. Of course these beliefs should not be completely arbitrary, they should respect the information structure of the game and they should be consistent with the equilibrium strategies whenever possible. Formally, a system of

beliefs is a mapping μ that assigns a probability distribution to the nodes in h for any information set h and a *sequential equilibrium* is defined as a pair (s, μ) consisting of a strategy vector s and a system of beliefs μ satisfying the following two conditions:

$$\begin{aligned} &s \text{ is } \textit{sequentially rational} \text{ given } \mu, \text{ i.e.} \\ &u^{\mu}_{ih}(s) \geq u^{\mu}_{ih}(s \backslash s'_i) \text{ for all } i, \text{ all } s'_i \in S_i, \text{ all } h \in H_i, \text{ and} \end{aligned} \tag{2.3}$$

$$\begin{aligned} &\mu \text{ is } \textit{consistent} \text{ with } s, \text{ i.e. there exists a sequence } s^k \\ &\text{of completely mixed behavior strategy vectors with } s^k \rightarrow s (k \rightarrow \infty) \\ &\text{such that } \mu(x) = \lim_{k \rightarrow \infty} p^{s^k}(x)/p^{s^k}(h) \text{ for each information set } h \\ &\text{and each } x \text{ in } h. \end{aligned} \tag{2.4}$$

(s^k is said to be completely mixed if $s^k_{ih}(c) > 0$ for all $i, h \in H_i$ and all choices at h). Condition (2.3) expresses that, given the beliefs μ, the player maximizes his payoff at h by playing according to s as long as the opponents play according to s as well. The consistency requirement means that, at an information set which a player does not expect to be reached, the beliefs can be explained by means of small trembles from the equilibrium strategies. (At other information sets the beliefs coincide with those induced by the equilibrium.) This consistency requirement is inspired by Selten's concept of trembling hand perfect equilibria (see the next section) but it is not completely intuitive on its own and Kreps/Wilson express some doubts about whether consistency actually "ought" to be defined as in (2.4). In fact Kreps and Wilson's intuitive motivation for where the beliefs come from does not involve any trembles. They argue that at information sets h that are initially assigned probability zero ($p^s(h) = 0$), it is plausible to assume that the player will construct some alternative hypothesis s' as to how the game has been played that is consistent with his observation (i.e. $p^{s'}(h) > 0$) and then use s' and Bayes' rule to compute his beliefs. Formally, define a system of beliefs to be *structurally consistent* if for each information set h there exists some s' with $p^{s'}(h) > 0$ and $\mu(x) = p^{s'}(x)/p^{s'}(h)$

for each x in h. (Kreps/Wilson then go on to strengthen this condition by requiring that alternative hypotheses at different information sets be related in certain ways.)

Although this structural consistency requirement seems intuitive at first, further reflection reveals that it actually is not. First, the idea of reassessing the game (i.e. to construct alternative hypotheses) runs contrary to the idea that a rational player can foresee and evaluate all contingencies in advance. (Recall the remarks on strategy vectors from the beginning of this section.) Secondly, structural consistency conflicts with the sequential rationality requirement (2.3). The latter requires believing that from h on play will be in accordance with s while the former requires believing that play has been in accordance with s'. Although these requirements are not conflicting in games with a stage structure (in these the past can be separated from the future) they may be incompatible in games in which the information sets cross, since in these deviations in the past are automatically accompanied by deviations in the future. An explicit example is contained in Kreps and Ramey (1987). That paper also contains an example of a game in which there does not exist a sequential equilibrium (s, μ) in which in addition μ is structurally consistent, hence, structural consistency may conflict with consistency. Since, as seen above, structural consistency does not seem a compelling requirement, one should not be bothered by this discrepancy. Of course there remains the question of whether the consistency requirement (2.4) can be expressed directly in terms of the basic data (i.e. the choices and information sets) of the extensive form of the game. The affirmative answer to this question is given in Kohlberg and Reny (1991).

The literature offers a variety of equilibrium concepts (usually under the common name of "perfect Bayesian equilibrium") that are related to the sequential equilibrium concept but in which milder restrictions are imposed on the way in which beliefs are formed in zero probability events. The weakest of these do not impose any conditions off the equilibrium path and allow, for example, that different players with "identical" information explain an unexpected deviation in different ways. (Note that (2.4), according to the usual "common knowledge" assumption underlying Nash equilibrium, assumes that

all players have a common theory to explain deviations.) For further details the reader is referred to Fudenberg and Tirole (1989) and Weibull (1990).

Other authors have proposed to impose additional requirements on the way beliefs are revised. In applications, such as the study of dynamic games with incomplete information, frequently the so-called support restriction is imposed. (For example, the concept of perfect sequential equilibrium (PSE, Grossman and Perry (1986)) that is often used in applications imposes this restriction.) This restriction requires that, if at a certain point in time a player assigns probability zero to a certain type of the opponent, then from that time on he continues to assign probability zero to that type. The restriction enables analysis by means of a dynamic programming procedure in which the beliefs are used as a state variable. However, Madrigal, Tan and Werlang (1987) have shown that imposing this restriction may lead to nonexistence: The support restriction may be incompatible with the (very mild) requirement that the beliefs be derived from the equilibrium strategies on the equilibrium path. The following example (taken from Nöldeke and Van Damme (1990)) demonstrates why this is the case and makes clear that the support restriction has nothing compelling to it. (For a more economic example, see Vincent (1990).)

[Insert Figure 2.a and 2.b here]

Consider the signalling game from Figure 2.a: Nature first selects a type of player 1, both possibilities being equally likely. If player 1 chooses L the game ends, otherwise player 2 has to choose between l and r. (Figure 2.b gives a convenient matrix representation of this game following the conventions outlined in Banks and Sobel (1987): The matrices correspond to the choices of player 1, the rows represent the types of this player and the columns are the choices of player 2.) The game has a unique Nash equilibrium,

viz. (L, R, r). Now consider the two-fold repetition of this game: Player 1's type is drawn once and for all at the beginning of the game, and before the beginning of round 2 only the actions from the previous round, but not the payoffs, are revealed. We claim that this game has a unique Nash equilibrium outcome, viz. type t_1 chooses L twice and type t_2 chooses R twice. (Proof: Strict dominance implies that type t_1 chooses L in both rounds and that type t_2 chooses R in the last round. Hence, type t_2 will choose LR or RR, or a mixture of these. LR cannot be type t_2's equilibrium strategy since (when player 2 plays his best response) it yields less than the payoff that type t_2 can guarantee himself by playing LL. Type t_2 cannot mix, since then player 2's unique best response is to choose r whenever R is chosen and this implies that RR is strictly better.) To support the unique equilibrium outcome, player 2 should choose r with a probability of at least $5/6$ in the second round after having observed L in the first round and R in the second. However, such behavior is not optimal if beliefs are required to be consistent with the equilibrium strategies as well as to satisfy the support restriction. Namely, these requirements force player 2 to believe that he is facing type t_1 for sure if he observes LR (since only t_1 chooses L in the first round in equilibrium) and if he has such beliefs he should play l. Hence, the beliefs associated with any Nash equilibrium necessarily violate the support restriction. The example makes clear that such a violation is actually quite natural: After having observed L in the first round, player 2 has no evidence that play is not in agreement with the equilibrium so he adopts equilibrium beliefs. After having observed LR, however, he has such evidence and he corrects his initial beliefs since after all it is only t_2 who might have had an incentive to try to mislead him.

3 Perturbed games

Selten (1975) proposes to escape from counterfactuals associated with irrational moves of rational players by giving up the assumption that players are perfectly rational and he introduces a model of slight imperfect rationality that is based on the idea that with some very small probability a player will make a mistake. He writes (Selten (1975, p. 35))

> "There cannot be any mistakes if the players are absolutely rational. Nevertheless a satisfactory interpretation of equilibrium points in extensive games seems to require that the possibility of mistakes is not completely excluded. This can be achieved by a point of view which looks at complete rationality as a limiting case of incomplete rationality."

Selten's approach is reviewed in subsection 3.1.

Alternatively, one may escape from the counterfactuals by giving up the assumption that the game fully describes the real situation. One may argue that the model is overabstracted, that there are always some aspects that are not incorporated and that, if a complete model were built, the difficulties associated with unreached information sets would vanish. That there are rewards associated with not taking the description of the game too literally is already known since Harsanyi (1973) in which it was shown that if the slight uncertainty that each player has about the payoffs of his opponents is actually taken into account, the usual instabilities (and interpretational difficulties) of mixed strategy equilibria vanish. (At least this holds for generic normal form games). In subsection 3.2 we briefly discuss several variants of the idea that a self-enforcing equilibrium should still make sense when the aspects that were abstracted away from (such as payoff uncertainty) are explicitly taken into account. It will turn out, that the results depend crucially on which story that one tells, and that even Nash equilibria that are not subgame perfect can make sense in certain contexts. Hence, a main conclusion to be drawn from subsection 3.2 is that, if the game model is not complete, it may not be appropriate to apply equilibrium refinements.

3.1 Perfect equilibria

In Selten (1975) incomplete rationality is modelled by the assumption that at each of his information sets a player will, with a small (independent) probability suffer from "momentary insanity" and make a mistake. Selten assumes that in the case of a mistake at the information set h the player's behavior at h is governed by some unspecified psychological mechanism which selects each choice at h with a strictly positive probability. Since

in such a perturbed game there are no unreached information sets, a Nash equilibrium prescribes the playing of a best response everywhere. Selten proposes to restrict attention to those equilibria of the original game that can be obtained as a limit of a sequence equilibria of perturbed games as the trembles vanish and he calls these perfect equilibria.

It is convenient to define perfect equilibria first for normal form games, i.e. games in which each player has to make just one choice and in which choices are simultaneous. Let $G = (S_i, u_i)_{i=1}^n$ be such a game, let σ be a completely mixed strategy vector (with σ_i representing the choice of player i if he makes a mistake) and let ε be a positive n-vector of mistake probabilities. Denote by $s^{\varepsilon,\sigma}$ the strategy vector that results if each player intends to play s and players make independent mistakes according to (ε, σ). (Hence, $s_i^{\varepsilon,\sigma}$ is the convex combination of s_i and σ_i that assigns weight ε_i to σ_i.) In the perturbed game $G^{\varepsilon,\sigma}$ (i.e. the game in which the players take the mistakes explicitly into account), the strategy vector s is an equilibrium if

$$u_i(s^{\varepsilon,\sigma} \backslash s_i) \geq u_i(s^{\varepsilon,\sigma} \backslash s_i') \quad \text{for all } i \text{ and all } s_i' \in S_i. \tag{3.1}$$

The strategy vector s is said to be a *perfect equilibrium* of G if s is a limit of a sequence $s(\varepsilon_n, \sigma_n)$ of equilibria of perturbed games $G^{\varepsilon_n,\sigma_n}$ with $\varepsilon_n \to 0$. Note that for s to be perfect it is sufficient to find *one* mistake sequence that justifies s. Selten (1975) proved that perfect equilibria exist and he showed that the strategy vector s is a perfect equilibrium if and only if s is a best response to a sequence of completely mixed strategy vectors that converges to s. In particular it follows that a perfect equilibrium is undominated (admissible).

Now let us return to an extensive form game Γ. Selten's assumption that trembles at different information sets are independent implies that one may think of different information sets of the same player as being administrated by different agents. The agent ih controlling the information set $h \in H_i$ has the same payoff as the original player i but this is the only link between agents, the agent ih cannot directly control

the actions of agent ik. Each agent maximizes for himself, counting on the rationality of the other agents, but incorporating the fact that they may make mistakes. It is now natural to look at the normal form game in which the agents are the players. This game $(S_{ih}, u_{ih})_{i=1,\, h\in H_i}^{n}$ (with, of course, $u_{ih} = u_i$ for all i and all $h \in H_i$) is called the *agent normal form* of Γ. A *perfect equilibrium* of the extensive for game Γ is defined as a perfect equilibrium of the agent normal form of Γ. Note that equilibria of a perturbed agent normal form game can be characterized by a condition similar to (3.1). This time we should satisfy the local condition

$$u_i(s^{\varepsilon,\sigma} \backslash s_{ih}) \geq u_i(s^{\varepsilon,\sigma} \backslash s'_{ih}) \text{ for all } i \text{ and all } h \in H_i, \tag{3.2}$$

where σ is a completely mixed behavior strategy vector. It is easy to see that each perfect equilibrium is a sequential equilibrium; Kreps and Wilson (1982a) proved that the converse holds for generic games.

A perfect equilibrium of the extensive form need not be perfect in the normal form. (Although this property does hold for generic extensive forms.) In Figure 3 the equilibrium (DL_1, L_2) is perfect in the extensive form: If player 1 fears that he is more likely to tremble than player 2 is, then his choice of D is optimal. The normal form assumes that each player can control his own actions completely. Obviously, in the normal form only (UL_1, L_2) is perfect. Note that in the normal form we represent the 'duplicate strategies' DL_1 and DR_1 by their 'equivalence class' D. This convention will be followed throughout the remainder of the paper. Hence, our normal form strategies will not specify what a player should do after he himself has deviated. The reader may fill in these actions in any way he wants without affecting the validity of any statements we make below about normal form strategies.

[Insert Figure 3 here]

The game from Figure 4 shows that, on the other hand, equilibria that are perfect in the normal form need not even be subgame perfect in the extensive form: (D_1, D_2) is perfect in the normal form since D_2 is player 2's best strategy if he believes that player 1 is more likely to tremble to A_1d than to A_1a. In the extensive form, perfectness excludes such beliefs: Even if player 1 trembled at his first move, player 2 should still consider it very likely that player 1 will play rationally (i.e. choose a) at his second move, hence, he should play A_2. Only (A_1a, A_2) is (subgame) perfect in the extensive form. (Note that the above conclusion would remain valid if the payoffs would be slightly perturbed so as to make the game generic. Reny (1988a) has shown that a normal form perfect equilibrium is always 'weakly sequential' in the extensive form.)

[Insert Figure 4 here]

Myerson (1978) argued that also in the normal form of Figure 4 it is nonsensical to believe that A_1d is more likely than A_1a. He argued that A_1d is a more costly mistake than A_1a, that a player will try harder to prevent more costly mistakes and that as a result these will occur much less often. Formally, he defined an ε-proper equilibrium of a normal form game as a completely mixed strategy vector s having the property that, if a pure strategy k of player i is a worse response against s than a pure strategy l, then the probability that s_i assigns to k is at most ε times the probability that s_i assigns to l. A limit of a sequence ε-proper equilibria (as ε tend to zero) is called a *proper equilibrium*. Such an equilibrium exists and is obviously perfect. An important property is that proper equilibria of a normal form game induce sequential equilibrium outcomes in every extensive form game with that normal form. Formally, if Γ is an extensive form game with normal form G and if s is a proper equilibrium of G, then there exists a sequential equilibrium (s', μ) of Γ such that $p^s = p^{s'}$. (Kohlberg and Mertens (1986),

Van Damme (1984)).

Another refinement that is related to the perfectness concept is the persistent equilibrium (Kalai and Samet (1984)). If $G = (S_i, u_i)_{i=1}^n$ is a normal form game and R_i is a compact convex subset of S_i for each i, then $R = X_i R_i$ is said to be an essential retract if there exists a neighborhood R' of R such that for each s' in R' there is some s in R that is a best reply against s'. (Roughly this definition strengthens perfectness by requiring stability against *all* perturbations; simultaneously it weakens perfectness by allowing sets of solutions, this in order to guarantee existence.) A minimal essential retract is called a persistent retract and an equilibrium that lies in such a retract is said to be a *persistent equilibrium*. Persistency does not seem to be a necessary requirement for self-enforcingness. For example, in the Battle of the Sexes Game of Figure 7.a only the pure equilibria are persistent, hence, a symmetric game need not have symmetric persistent equilibrium. Similarly, in the coordination problem of Figure 5 the outcome in which player 1 chooses D seems perfectly self-enforcing if players cannot communicate. (Note that player 2 has no incentive whatever to communicate.) However, only the two equilibria with payoff (3,3) are persistent in this game. From these examples it appears that persistency is more relevant in an evolutionary or in a learning context, rather than in a pure eductive context.

[Insert Figure 5 here]

3.2 Correlated Trembles

Selten's assumption that mistakes are uncorrelated across different information sets has been criticized and it has been argued (for example, in Binmore (1987)) that in some contexts it may be more natural to allow correlated trembles. Obviously, if perturbations in a more general class are allowed and if only stability against one sequence of perturbed games is required, then typically less outcomes will be eliminated. Correlated trembles arise naturally if there is initial uncertainty about the payoffs and we will now

give some examples to illustrate that less equilibria can be eliminated in this context, in fact, that in some cases no Nash equilibrium can be eliminated. The reason is that, when there is initial payoff uncertainty, the players beliefs may change drastically during the game. Possibilities which are unlikely ex ante may have large effects ex post when they actually happen. Consequently, it is by no means obvious that the perturbations like the ones discussed below should be considered *slight* perturbations. (That a small amount of payoff uncertainty may have a large effect is also known from the 'applications' in Kreps and Wilson (1982b), Kreps et al. (1982) and Fudenberg and Maskin (1986). The results below are different since they show that even vanishing uncertainty may have drastic consequences.)

Consider once more the game $\Gamma_2(2)$ from Figure 1.b but suppose now that player 1 initially has some doubts about the objectives of player 2. He believes that with probability $1-\varepsilon$ player 2 is 'rational' and has payoffs as in $\Gamma_2(2)$ and that with probability ε this player is 'irrational' and tries to minimize player 1's payoffs (hence, in this case $u_2 \equiv -u_1$). Player 2 knows his own objectives. The subgame perfect equilibrium (A, D_2a) of the original game is no longer viable in this context: If player 2 believes that player 1 chooses A, then he is facing the irrational type of player 2, hence, player 1 should deviate to D. The reader easily verifies that the perturbed game has a unique subgame perfect equilibrium and that in this equilibrium player 1 chooses both A and D with probability $1/2$ while the rational type of player 2 chooses A_2 with probability $2\varepsilon/(1-\varepsilon)$. Hence, with this story, although we obtain the subgame perfect equilibrium outcome of the game $\Gamma_2(2)$ in the limit, we rationalize a strategy for player 1 that is not this player's subgame perfect equilibrium strategy.

By using a construction as above, Fudenberg et al. (1988, Proposition 4) have shown that, for every extensive form game, each equilibrium that is (strictly) perfect in the normal form can be similarly rationalized by a sequence of slightly perturbed games in which each player has some private, independent, information about his own payoffs. Hence, also outcomes that are not subgame perfect can be 'rationalized' by means of

slight payoff uncertainty. This result can be illustrated by means of the extensive form game of Figure 4 which has (A_1a, A_2) as its unique subgame perfect equilibrium outcome. Suppose that player 2 believes that with a small but positive probability player 1 has the payoff 4 if D_2 or d is played. (All other payoffs remain as in Figure 4 and it is assumed that player 1 knows which payoffs prevail.) This perturbed game has a strict Nash equilibrium (i.e. each agent chooses his unique best response) in which player 2 chooses D_2 while player 1 chooses D_1 if his payoffs are as in Figure 4. In this equilibrium player 2 correctly infers from the choice of A_1 at player 1's first information set that this player will choose d at this second move, this induces him to choose D_2 which in turn makes D_1 strictly optimal for the 'regular type' of player 1. Hence, in the limit, as the uncertainty vanishes we obtain the (normal form perfect) equilibrium (D_1, D_2).

Fudenberg et al. (1988) also show that if the information of different players may be correlated one can rationalize the larger set of normal form "correlated perfect" equilibria, and that, if it is possible that some player i may have information about the payoffs of player j that is superior to j's information, then one may even rationalize the entire set of pure strategy Nash equilibria. (Formally, if s is a pure strategy Nash equilibrium of game Γ then there exists a sequence of slightly perturbed games in which each player has some private information and an associated sequence of strict equilibria that converges to s (Fudenberg et al. (1988, Proposition 3)). This result may be illustrated by means of the game of Figure 1.a. Assume that with a small probability ε the payoffs associated with (A, d) are (2,2) rather than (0,0) and that only player 1 knows what the actual payoffs are. In this perturbed game it makes perfectly good sense for player 1 to choose D if the payoffs are as in Figure 1.a, since he may fear that player 2 may interpret the choice of A as a signal that the payoffs are (2,2) and continue with d after A.

The driving force behind the previous example was that the players could correlate their strategies. Of course, payoff uncertainty is not necessary for correlation to be possible (Aumann (1974)) and, consequently, constructions like the above may be possible if the payoffs are known but there is uncertainty about the game structure. As an example

consider the game from Figure 6 in which the simultaneous move subgame is played between the players 1 and 2 (player 3 is a dummy in that game). Since the subgame has a unique Nash equilibrium with value (3,3,4), the unique subgame perfect equilibrium yields each player the payoff 5. Note, however, that the subgame also admits a correlated equilibrium c in which each of the nonzero entries is played with probability $1/6$. If player 3 believes ex ante that there is an ε probability that the players 1 and 2 have a correlation device available that enables them to play c, then he can interpret the choice of A_1 as a signal that this device is available. This interpretation leads him to choose A_3 which in turn induces player 1 to choose D_1 if the correlation device is not available. Hence, this story justifies the outcome (4,4,8). (For an elaboration on this example, and an in-depth study of 'sequential correlated equilibria' the reader is urged to consult Myerson (1986) and Forges (1986).)

[Insert Figure 6 here]

The point of this subsection has been to show that if the game model is incomplete, then one cannot tell which equilibria are self-enforcing without knowing where the incompleteness of the model consists of, i.e. without knowing the context in which the game is played. Consequently, in the next section we return to the classical point of view that

> "the game under consideration fully describes the real situation, — that any (pre)commitment possibilities, any repetitive aspect, any probabilities of error, or any possibility of jointly observing some random event, have already been modelled in the game tree" (Kohlberg and Mertens (1986, Fn. 3).

4 Stable equilibria

The game of Figure 7.b shows that the concepts introduced thus far do not provide sufficient conditions for self-enforcing equilibria. In this game player 1 has to choose between an outside option yielding both players the payoff of 2 or to play the Battle of the Sexes games from Figure 7.a. One equilibrium has player 1 choosing D while the players continue with (w, s) if the BS-subgame is reached. The equilibrium is perfect since perfectness allows player 2 to interpret the move A of player 1 as an unintended mistake which does not affect player 1's behavior at his second move. However, there clearly exists a much more convincing explanation for why the deviation occurred. Player 2 should realize that player 1 (being a rational player) will never play Aw since this is strictly dominated by D. Hence, he should conclude that the deviation signals that player 1 intends to play s in the subgame and he should respond by playing w. Clearly, this chain of reasoning upsets the equilibrium.

[Insert Figures 7.a and 7.b here]

Note that the above argument involved the normal form of the game, we discussed strategies for the entire game rather than independent actions at different information sets. (Formally, the argument amounts to the observation that, by eliminating dominated strategies, one can reduce the game to the outcome (As, w). Hence, the example shows that perfect equilibria are not robust to the elimination of dominated strategies.) What is involved is an argument of Forward Induction: Player 2's beliefs and actions should not only be consistent with deductions based on player 1's rational behavior in the future (this is the sequential rationality requirement captured by perfectness) but they should also incorporate (at least as much as possible) rational behavior of player 1 in the past. It seems that the latter type of considerations can only be incorporated by taking a global picture, i.e. by looking at the normal form.

Kohlberg and Mertens (1986) argue forcefully (and convincingly) that the normal form of a game contains sufficient information to find the self-enforcing equilibria of this game. The argument is simply that rational players can and should always fully anticipate what they would do in every contingency; a theory of rationality that would tell a player at the beginning of the game to choose c if the information set h *were to* be reached and that would simultaneously advise the player to take a different action c' if h *is* actually reached is hardly conceivable. This classical point of view implies that self-enforcing equilibria can only depend on the normal form of the game and entails that (subgame) perfect and sequential equilibria are unsatisfactory. (The two games in Figure 4 have the same normal form but they have different sets of perfect (resp. sequential) equilibria. Note that in a normal form game every Nash equilibrium is sequential.)

Kohlberg and Mertens argue further that one even needs less information than is contained in the normal form to find the self-enforcing equilibria: Since players are always explicitly allowed to randomize over pure strategies, adding such mixtures explicitly as pure strategies in the game should not change the solutions. This requirement implies that the solutions of a game can only depend on the so-called reduced normal form, i.e. on the normal form that results when all pure strategies that are convex combinations of other pure strategies have been deleted. It turns out that this invariance requirement is incompatible with the requirement that the solution of an extensive form game be a subgame perfect equilibrium: There exist two games with the same reduced normal form that have disjoint sets of subgame perfect equilibria. An illustration is provided by the game of Figure 8.a. This is the (reduced) normal form of an extensive form game in which player 1 chooses between an outside option l yielding 2 or to play a 2×2 subgame of the matching pennies type, and the unique subgame perfect equilibrium of this game has player 2 choosing $\frac{1}{2}L + \frac{1}{2}R$. If we add the mixture $s = \frac{1}{2}l + \frac{1}{2}m$ explicitly as a pure strategy of player 1 then we obtain the normal form from Figure 8.b which corresponds to an extensive form game in which player 1 chooses between an outside option or to play a 3×2 subgame (with strategies s, m and r for player 1 and L and R for player 2).

The reduced normal form of this game is as in Figure 8.a, however, the unique subgame perfect equilibrium of the game requires player 2 to choose R.

[Insert Figures 8.a and 8.b here]

The incompatibility of the two requirements calls again into question of whether subgame perfectness is really necessary for self-enforcingness. Like the examples in Section 2, this example suggests that one adopts a more liberal point of view and allows multiple beliefs and multiple recommendations for player 2. There is certainly no need to specify a unique action for this player since his choice doesn't matter anyway when he plays against a rational opponent. His choice may matter if his opponent plays irrationally but then the optimal choice probably depends on the way in which player 1 is irrational and since no theory of irrationality is provided, the analyst should be content to remain silent. Generalizing from this example one might argue that we may be satisfied if we can identify the outcomes resulting from rational play, i.e. if we can specify which actions a player should take as long as the opponents' behavior does not contradict their rationality. A self-enforcing norm of behavior should not necessarily pin down the players' behavior and beliefs in those instances which cannot be observed when the norm is in effect.

Kohlberg and Mertens also argue that, besides failing to satisfy invariance, a second reason for why perfect (and sequential) equilibria are not satisfactory concepts is that they may allow equilibria in dominated strategies. (Perfectness implies that all moves are undominated, however, the overall strategy may be dominated, cf. the equilibrium (DL_1, L_2) in Figure 3.) Kohlberg and Mertens consider admissibility of the equilibrium strategies (i.e. these strategies not being weakly dominated) to be a fundamental requirement. Furthermore, as we have seen when discussing the game from Figure 7.b, yet another drawback of perfect (and sequential) equilibria is that they are not robust to

the (iterative) elimination of dominated strategies. Kohlberg/Mertens argue that since (weakly) dominated strategies are never actually chosen by rational players and since all players know this, such strategies can have no impact on whether or not an equilibrium is self-enforcing. This requirement of "independence of dominated strategies" again points to a set-valued solution concept, since, as is well-known, the outcome of the elimination process may depend on the order in which the strategies are eliminated. For example, in the game of Figure 8.a, the elimination order m, R, r leads to the conclusion that player 2 should play L, while the order r, L, m leads to the conclusion that he should play R. Again one sees that multiplicity is natural: If player 2 eliminates a dominated strategy of player 1 he attributes rationality to this player, but he may have to move only if player 1 actually is irrational. We simply reconfirm that the way in which player 1 is irrational determines player 2's choice and that, if one does not specify what irrational behavior looks like, one should not necessarily specify a unique choice for player 2.

[insert Figure 9 here]

A more interesting example, in which different elimination orders actually produce different outcomes is provided by the game from Figure 9. In this game the notions of forward and backward induction are conflicting. Backward induction (or the elimination order al, AL, d, AR) leads to the conclusion that player 1 should choose D and that the payoffs will be (2,0). Forward induction, or more precisely the fact that player 2 interprets the choice of A as a signal that player 1 will not play R, yields as a possible elimination order AR, ar, D, di, which gives the conclusion that player 1 should play AL and that player 2 should choose d resulting in the payoffs (2,2). (This game is nongeneric since both D and d yield player 1 the payoff 2, however note that, when one does the backward induction, there are never ties.) This example shows that, if we indeed insist on the requirement that self-enforcing norms should be "independent of dominated strategies"

then, in nongeneric games, we cannot identify norms with outcomes and this raises the question of how to define norms in this case. Kohlberg and Mertens show that the set of Nash equilibria of a game consists of finitely many connected components and they suggest as candidates for self-enforcing norms (connected subsets of) such components. Since generic extensive form games have only finitely many Nash equilibrium outcomes (Kreps and Wilson (1982a)) it follows that for generic games all equilibria in the same component induce the same outcome, so that for such games the Kohlberg/Mertens suggestion is only a relatively minor departure from the traditional notion of a single-valued solution.

The requirement that the solution be "independent from dominated strategies" is a global requirement: Strategies that are 'bad' from an overall point of view will not be chosen, hence, they should play no role. Once a specific norm is under consideration one can be more specific. If the norm is really self-enforcing then a player will certainly not choose a strategy that, as long as the others obey the norm, yields him strictly less than he gets by obeying the norm. Hence, for a norm to be self-enforcing it is necessary that it remains self-enforcing after a strategy has been eliminated that is not a best reply against the norm. The power of this requirement of "independence of non-best responses" (INBR) will be illustrated in the next section. The game $\Gamma_2(2)$ from Figure 1 shows that this INBR requirement is not satisfied by the subgame perfect equilibrium concept: strategy A_2a of player 2 is not a best response against player 1's equilibrium strategy A, but if A_2a is deleted from the game, player 1 will switch to D. Hence, if one wants to satisfy INBR as well as some form of sequential rationality one is again forced to accept a set-valued solution concept.

Having specified several necessary conditions for self-enforcingness, the obvious question, of course, is whether it is possible to satisfy all these requirements. The answer is yes: There exist norms satisfying the properties discussed above as well as some other desirable properties.

Theorem: *(Mertens (1988, 1989a, 1990).) There exists a correspondence that assigns to each game a collection of so-called stable sets of equilibria such that*

(i) (connexity and admissibility) each stable set is a connected set of normal form perfect (hence, undominated) equilibria.

(ii) (invariance) stable sets depend only on the reduced normal form.

(iii) (backward induction) each stable set contains a proper (hence, sequential) equilibrium.

(iv) (iterated dominance) each stable set contains a stable set of a game obtained by deleting a (weakly) dominated strategy.

(v) (INBR) each stable set contains a stable set of a game obtained by deleting a strategy that is not a best response against any element in the set.

(vi) (player splitting property) stable sets do not change when a player is split into two agents provided that there is no path in the game tree in which the agents act after each other.

(vii) (small worlds property) If there exists a subset N' of the player set N such that the payoffs to the players in N' only depend on the actions of the players in N', then the stable sets of the game between the players in N' are exactly the projections of the stable sets of the larger game.

The properties (i) – (v) have already been discussed. Property (vi) implies that it does not matter whether a signalling game (see the next section) is analyzed in normal form (2 players) or in agent normal form, or in any intermediate game form. Note that this property does not hold if two agents of the same player move after each other: The outcome (2,2) is stable in the agent normal form of the game of Figure 7.b: If player 1 consists of two separate agents then the first has no control over the second and he cannot signal this agent's intentions. Property (vii) is a decomposition property that guarantees that the solutions of a game do not depend on things that have nothing to do with the game. Note that we naturally have 'contains' rather than 'is' in (iv) and (v): stable sets may shrink if 'inferior' strategies are deleted. Intuitively, stable sets have to be large since they must incorporate the possibility of irrational play (and there seems no unique best

way to play against irrational opponents); however, by eliminating dominated strategies one attributes more rationality to the players, makes them more predictable and this leads to a smaller set of optimal actions, hence, to smaller stable sets. The game $\Gamma_2(2)$ of Figure 1 provides an illustration. The set of normal form perfect equilibria of this game consists of the strategy vectors $s = (s_1, s_2)$ with $s_1 = pA + (1-p)D$, $s_2 = D_2$ and $p \leq 1/2$, hence, (by (i)) each stable set is a subset of this set. Let S^* be a stable set. Since the strategy A_2d is not a best response against S^* and since, in the game in which A_2d is deleted, the unique stable set is (A, D_2) (by admissibility), we have that (A, D_2) belongs to S^*. Similarly the strategy $1/2A + 1/2D$ of player 1 must belong to S^*, for, if this would not be the case, then A_2a would be 'inferior' so that (by (i) and (v)) (D, A_2d) should belong to S^* but this is impossible. Hence, it follows by (i) that in $\Gamma_2(2)$ the unique stable set is the set of all normal form perfect equilibria.

Note that the Theorem is stated as an existence theorem, it does not say how to find stable sets. Kohlberg and Mertens (1986) initially defined a stable set of a game G as a "minimal closed set S of equilibria of G with the property that each perturbed game $G^{\varepsilon,\sigma}$ (see Section 4) with sufficiently small ε has an equilibrium close to S". This definition is essentially the same as that for perfect equilibria except that one works with the normal form and that one has to look at *all* perturbations rather than just one sequence. However, it turned out that this concept failed to satisfy some essential properties from the Theorem (such as (iii)). Mertens (1988, 1989a) refined the definition to remedy this deficiency and proved the Theorem. For the purpose of this paper the exact definition is not so relevant, since in the applications to be discussed next, the properties from the Theorem will suffice to single out the stable outcomes. Finally, Hillas (1990) defines a stable set of a game G as "a minimal closed set S of equilibria of G with the property that for each game G' with the same reduced normal form as G and for each upper-hemicontinuous compact convex valued correspondence that is pointwise close to the best reply correspondence of G' there exists a fixed point that is closed to S". Such stable sets exist and satisfy the properties (i) – (v) from the Theorem.

5 Forward induction

In this section we briefly discuss some applications of forward induction, i.e. of the idea that the inferences players draw about a player's future behavior should be consistent with rational behavior of this player in the past. Informally stated, forward induction amounts to the requirement that for an equilibrium to be self-enforcing there should not exist a nonambiguous deviation from the equilibrium that, when interpreted in the appropriate way, makes the deviator better off. This attractive idea has proved elusive and, consequently, several formalizations have been proposed in the literature. It has turned out, however, that stability (and in particular "independence of dominated strategies and/or non-best responses") captures at least some of the forward induction logic. In this section we first illustrate some applications of stability in games of complete information, thereafter, we indicate how powerful that concept is to eliminate implausible equilibria in signalling games. Along the way several other formalizations of forward induction that are in some way related to stability will be encountered. Throughout the section attention will be confined to generic games, i.e. to games that have finitely many Nash equilibrium outcomes. We will call an outcome of such a game stable if there exists a stable set of which all elements induce this outcome. (Recall that in generic games all elements in a same stable set yield the same outcome.)

5.1 Signalling intentions

Consider the following modification of the game of Figure 7.b: First chance determines whether player 1 or player 2 will have an outside option available. If a player takes up the outside option each player has the payoff 2. If player i is selected by chance but he does not take his option then players play the Battle of the Sexes. It is easily seen that in the unique stable outcome the option is not taken up, that the player who has the option available chooses to play BS and gets the payoff 3. (Abdalla et al. (1989) provide experimental evidence on the success of forward induction in similar games). In particular, we see that the history of the game determines the way in which the subgame is played: The player's expectations in the subgame are not endogenous i.e. they are

not determined by the subgame alone but depend on the context in which the subgame arises. (See Mertens (1989b) for an informal discussion on this topic.) The equilibrium selection theory of Harsanyi and Selten (1988) is based on the assumption of endogenous expectations: Harsanyi and Selten impose the requirement of subgame consistency, i.e. a subgame should always be played in the same way no matter how it arose. The example shows that subgame consistency conflicts with stability. Similarly, it may be shown that also other concepts that require history independence, such as Markov perfection (Maskin and Tirole (1989)) or stationarity conflict with stability.

Suppose that the players have to play the Battle of the Sexes Game from Figure 7.a but that before playing this game player 1 has the option of burning one unit of utility and that when BS is played it is common knowledge whether or not player 1 burned utility. It is easily seen that iterative elimination of dominated strategies reduces the normal form to the payoff (3,1), hence, only the outcome in which player 1 does not burn utility and gets his most preferred outcome is stable. Using this argument, Ben-Porath and Dekel (1987) have shown that in games of "mutual interest", the players will succeed in coordinating on the Pareto best equilibrium if one player has the ability to destroy utility. In Van Damme (1989) it is shown that 'in the Battle of the Sexes' all stable outcomes are inefficient (i.e. involve some burning) if both players have the opportunity to simultaneously burn utility. Applications of these ideas to more economic contexts are found in Bagwell and Ramey (1990), Dekel (1989) and Glazer and Weiss (1990).

The deletion of dominated strategies in the BS with one-sided burning of utility corresponds to the following intuitive story: If player 2 observes that player 1 burns utility he should conclude that player 1 will continue with s; assuming that player 1 will play w does not make sense since burning followed by w yields at most the payoff zero, hence, is strictly dominated by not burning and randomizing between s and w. This conclusion leads player 2 to play w if burning is observed and burning utility is sure to yield player 1 the payoff 2. At this stage of the reasoning process we are back to a game like that in Figure 7.b and we can continue reasoning as in that example to reach the conclusion

that (s, w) should also be played if player 1 does not burn utility. A little reflection reveals that the argument above is not intuitive at all: It is not clear why player 2 should respond to the burning by playing w since, given the conclusion we just reached, burning is a signal that player 1 is not rational, at least it signals that he did not follow the above reasoning. At this point the reader should be reminded of the discussion of counterfactuals in Section 2, so it is not necessary to go into details here. Let us just remark that stability does not force player 2 to play w after player 1 has burned utility: The stable set includes both ww and ws for player 2 ($\alpha\beta$ denotes that player 2 responds to not burning by α and to burning by β). Namely, property (iv) of the Theorem implies that $(-s, ww)$ belongs to the stable set. ($-s$ denotes the strategy of not burning and playing s.) Furthermore, given that player 2 plays a mixture of ww and ws in any element of the stable set, the strategies bs (i.e. burning and then playing s) and $-w$ are inferior for player 1. If these strategies are eliminated, ws becomes dominated for player 2 and the normal form is reduced to $(-s, ws)$, so that the Theorem implies that this strategy pair also has to belong to the stable set.

When a game with multiple equilibria is repeated the set of subgame perfect equilibrium payoffs expands until in the limit it covers, at least under a mild regularity condition, the entire set of feasible and individually rational payoff vectors. This is the content of the "Folk Theorem" (Benoit and Krishna (1985)). Hence, in repeated games the problem of multiplicity of equilibria is ubiquitous. Considerations of forward induction may eliminate some of these equilibria as the twice repeated battle of the sexes may show. As an illustration, let us show that the outcome (path) in which the one-shot equilibrium (s, w) is played twice is not stable. Namely, INBR implies that player 1 should interpret a deviation of player 2 to s in the first round as a signal that player 2 will also play s in the second round. (If he would plan to play w then his payoff is at most 1, which is less than the equilibrium payoff, hence, such a strategy is not a best response.) Consequently, after the deviation player 1 should play w but then player 2 gains by deviating (his payoff is 3 rather than 2), so that the outcome is not stable. Stable outcomes are alternating between (1,3) and (3,1), as well as playing the mixed

equilibrium twice, and some other mixtures in which the continuation at time 2 depends on the outcome of stage 1. It seems that stability forces payoffs to move closer to the 45° line but whether this property remains for repetitions with longer duration remains to be investigated.

To this author's knowledge no general results are available for stable equilibrium payoffs of repeated games: mathematically stability is not very easy to work with. Some preliminary results on repeated coordination games are contained in Osborne (1990). In particular, Osborne shows that in a class of repeated coordination games, paths that consist of pure Nash equilibria of the stage game can be stable only if they yield payoffs that are nearly Pareto optimal. This restriction on paths is unfortunate since for more general games no such path need be stable (Van Damme (1989)). Osborne does not use the full power of stability, he works with a weaker criterion of "immunity to a convincing deviation" (which is akin to the Cho and Kreps (1987) intuitive criterion (see the next subsection) and to the formalization of forward induction proposed in Cho (1987)). One negative result that is known is that stability conflicts with ideas of renegotiation-proofness: there may not exist a stable equilibrium that is also renegotiation-proof (Van Damme (1988)). (Renegotiation-proofness requires that at each stage of the game players continue with an equilibrium that is Pareto efficient within the set of the available equilibria, see Pearce (1990) for an overview of the various concepts formalizing this idea). In an interesting application Ponssard (1990b) shows that forward induction leads to the conclusion that long term competition in a market with increasing returns to scale forces firms to use average cost pricing. Ponssard, however, develops his own concept of forward induction (also see Ponssard (1990a, c)) and it is not clear that stable equilibria satisfy Ponssard's conditions.

An alternative (preliminary) formulation of forward induction based on an idea originally developed in McLennan (1985) was proposed in Van Damme (1989). In that paper it was argued that, in a generic 2-player game in which player 1 has the choice between an outside option o or to play a subgame γ of which a unique viable (say stable) equilib-

rium e yields player 1 more than his option, only the outcome in which player 1 chooses to play γ and e is played in γ is sensible. The justification for this requirement is that by choosing to play γ player 1 can unambiguously signal that he will play according to e in γ. Alternatively one may imagine a context in which there is initial strategic uncertainty about whether the norm o or the norm γe is in effect: Even if player 2 originally believes that he is in a world in which o is obeyed, he concludes from the fact that he has to move that the norm must be γe and he responds appropriately. (Telling the story in this way makes clear that this type of forward induction is related to the risk dominance concept from Harsanyi and Selten (1988). Another paper dealing with this type of situations is Suehiro (1990). Also Binmore (1987) has such a context in mind when he presents an argument in favor of the imperfect equilibrium in Selten's 'horse' game.) Van Damme (1989) constructs an example to show that stable outcomes as originally defined by Kohlberg and Mertens do not necessarily conform to this forward induction logic. It is unknown to this author whether Mertens' refined stability concept satisfies this forward induction requirement.

5.2 Signalling private information

A signalling game is a 2-player game in which player 1, who has private information takes an action ('sends a signal') that is observable to player 2 who thereupon takes an action and in which the payoffs depend on both players' actions and the type (i.e. the information) of player 1. (Formally, a signalling game is a tuple $\Gamma = (T, M, (R_m)_m, u_1, u_2, \pi)$ where T is the (finite) set of types of player 1, M is the (finite) set of messages that can be send, R_m is the (finite) set of responses to $m, u_i = u_i(t, m, r)$ is the payoff function of player i, and π is a probability distribution on T representing the initial beliefs of player 2. An example of a signalling game is the game in Figure 2.a; from now on we will use a matrix representation as in Figure 2.b. to depict signalling games). Signalling games were introduced by Spence (1974) and they provide stylized models of many interesting economic situations (see Cho and Kreps (1987) and Kreps and Sobel (1991)). These games typically have large numbers of equilibria and researchers have used intuitive, context dependent arguments to eliminate equilibria. Although a great variety

of refinements exist, they all incorporate some form of forward induction, hence, they can be related to the stability concept from the previous section. Next we briefly discuss these relations. (The reader is referred to Cho and Kreps (1987), Banks and Sobel (1987), Kreps and Sobel (1991) and Sobel et al. (1991) for more details.) Before starting to discuss the relationships it should however be noted that the "intuitive criteria" are based on a somewhat different point of view, viz. economists have tried to directly define "plausible beliefs" and proposed to restrict attention to the ("plausible") equilibria that can be supported by "plausible beliefs". Such a requirement is stronger than the ones considered previously which were based on the idea that a candidate equilibrium should be rejected if it can be upset by "plausible" beliefs. The difference is that there may not exist equilibria that can be sustained by "plausible" beliefs since "plausible" beliefs may not exist (cf. the discussion on burning utility in the battle of the sexes game).

Let an equilibrium s of a signalling game be given. Typically the intuitive criteria that are used to judge the "plausibility" of this equilibrium start out by assuming that, if player 1 does not deviate from his equilibrium strategy, player 2 will not deviate either, hence, that playing the equilibrium strategy guarantees each type of player 1 his equilibrium payoff. (This assumption certainly makes sense: If the equilibrium is really self-enforcing, then no player will deviate. However, see the discussion in the Figures 12 and 13.) Next, assume that m is a message that is not sent if s is played. If choosing m is sure to yield a certain type t of player 1 less than what the equilibrium guarantees this type, then it is not "plausible" to assume that t will choose m and it should be possible to sustain the original equilibrium by beliefs that assign zero weight to t. Depending on how one defines "to sustain" in the previous sentence, the resulting test is known as *"the intuitive criterion"* or as *"equilibrium dominance"* (Cho and Kreps (1987)). The equilibrium s satisfies the intuitive criterion if for each type t of player 1 there exists a belief in the restricted set (of beliefs that puts zero weight on the types for which m is dominated) and an associated best response for player 2 at m that makes type t prefer to choose s rather than m. The test posed by equilibrium dominance is more restrictive and requires that there exists a belief in the restricted set and an associated best

response of player 2 such that no type of player 1 wants to deviate to m if that response is taken at m. Hence, the latter test requires that different types conjecture the same response after m, the former allows different types to have different conjectures. In the signalling game of Figure 10 the outcome in which both types of player 1 choose L does not survive application of the intuitive criterion since the latter requires that, after R, player 2 should put weight 1 on type t_1 and play l. In the game of Figure 11, the outcome in which all types choose L survives the intuitive criterion (this requires that player 2 puts weight zero on t_3 but it allows that the conjectures of t_1 and t_2 are mismatched, i.e. that t_1 believes that player 2 will play m and that t_2 believes that he will play l), but it does not pass the equilibrium dominance test, since if t_1 and t_2 conjecture the same (mixed) strategy of player 2, at least one of them will deviate. (Note that the game of Figure 11 (with t_3 deleted) demonstrates the claim made at the beginning of Section 2 that the Nash equilibrium concept depends in an essential way on the assumption that different players (here t_1 and t_2) conjecture the same out-of-equilibrium responses.)

[Insert Figures 10 and 11 here]

The above tests may be applied repeatedly. Formally, this repeated procedure runs as follows. Given an equilibrium s and an unsent message m, one first constructs the auxiliary signalling game in which player 1 has the choices s and m, where s guarantees the equilibrium payoffs and where the payoffs after m are the same as those in the original game. Next one starts eliminating strictly dominated strategies in the agent normal form of this game (hence, the types of player 1 are considered as independent players). If during the process the action s vanishes for some type t, then s does not satisfy the intuitive criterion. If the game that one obtains at the end of process does not have s as an equilibrium, then s fails the equilibrium dominance test. Since only actions are eliminated that are not a best response against any equilibrium in the same component as s, the Theorem implies that equilibria failing any of these tests cannot

belong to stable sets.

[Insert Figure 12 here]

Alternatively, one might construct the normal form of the auxiliary signalling game and eliminate dominated strategies in that game form. This poses a stricter test since more dominance relationships exist in the normal form. Consider the equilibrium s of the 3-message signalling game of Figure 12 in which both types choose L and the auxiliary game corresponding to the message M. (Hence, for the moment we completely neglect the message R.) Then s survives the equilibrium dominance test since choosing M is not dominated for either type. In the normal form, however, the strategy LM (i.e. t_1 chooses L and t_2 chooses M) is dominated (by a combination of ML and MM) and after this strategy has been eliminated one sees that player 2 should play l, thereby upsetting s. The intuitive argument corresponding to the elimination of dominated strategies in the normal form is known in the literature under the name of *co-divinity* (Sobel et al. (1991)), a criterion that is slightly weaker than that of *divinity* (Banks and Sobel (1987)). These criteria may also be described as follows. Assume that (the types of) player 1 conjecture that player 2 will reply to m with the response r. Letting $u^s(t)$ denote the equilibrium payoff of type t, the propensity $\lambda(t,r)$ for type t to deviate from s is given by

$$\lambda(t,r) = \begin{cases} 0 & \text{if } u^s(t) > u(t,m,r) \\ \in [0,1] & \text{if } u^s(t) = u(t,m,r) \\ 1 & \text{if } u^s(t) < u(t,m,r) \end{cases} \tag{5.1}$$

Hence, if player 2 knows that player 1 conjectures that he will play r, then his beliefs will be in the set

$$\mathcal{B}(\pi, r) = \{\pi' \in \Delta(T);\ \pi' = \pi\lambda(\cdot, r) \quad \text{for some } \lambda \text{ as in (5.1)}\}. \tag{5.2}$$

If there exists a possible conjecture r for which $\mathcal{B}(\pi, r)$ is not empty (i.e. if there exists a type that would not lose from deviating to m), then divinity and co-divinity require that the equilibrium s can be sustained by beliefs that belong to $\cup_r \mathcal{B}(\pi, r)$ where r ranges over the possible conjectures. Divinity is a slightly stronger concept since it allows only conjectures r that are (mixed) best responses while co-divinity allows the larger set of all mixtures of (pure) best responses. Banks and Sobel (1987) show that every stable component contains a divine equilibrium.

It will be clear that, because of (5.1), the divinity concepts force the updating to be monotonic: If type t_1 has a "greater incentive to deviate" to m than type t_2 has, then player 2 should not revise downward the probability that he is dealing with t_1 after m has been chosen. For example, in the game of Figure 12 both t_1 and t_2 could possibly gain by deviating from L to M but t_1 has the "greater incentive" to do so (the range of responses where t_1 gains is strictly larger than the range where t_2 gains) so that co-divinity requires that the posterior probability of t_1 after M is at least $1/2$; hence, player 2 should choose l thereby upsetting the equilibrium.

Note that divinity investigates each unsent message separately. (For each such message a separate auxiliary game is constructed, and s is eliminated if it fails the test in at least one auxiliary game.) In Figure 12, for example, it is thus required that player 2 plays l after M and l' after R. If player 1 foresees this reaction and plays his best response (R if t_1 and M if t_2) beliefs are induced that are incompatible with those of divinity. In fact, player 2's best response against this best response (viz. playing r after M and r' after R) sustains the original equilibrium. (Formally what happens is that, by including the third message, LM becomes undominated in the normal form.) Some readers might conclude from this that divinity is not an intuitive requirement after all. In the author's opinion the above argument simply shows that we do not know what will happen when the pooling equilibrium at L is recommended. However, this should not

bother us: we also do not know what will happen if, in an ordinary normal form game, a strategy vector is recommended that is not a Nash equilibrium. Questions concerning "disequilibrium dynamics", i.e. questions dealing with what will happen when a non-self-enforcing equilibrium is proposed, cannot be answered by equilibrium analysis. (Cf. Von Neumann and Morgenstern (1948, Section 4.8.2.).)

The so-called "Stiglitz critique" (Cho and Kreps (1987, p. 203)) on the intuitive criterion (or more precisely on the assumption that not deviating guarantees the equilibrium payoff) also involves such "disequilibrium dynamics". The critique may be illustrated by means of the game of Figure 13. In one equilibrium of this game, the types of player 1 pool at L and player 2 responds to L with l. The intuitive criterion eliminates this equilibrium: Type t_1 will deviate to R since he foresees that player 2 will switch to l' at R. According to the critique one should not stop the analysis with this disequilibrium outcome. Rather player 2 should realize that only t_2 can have chosen L and he should switch to r after L. But then t_2 also finds it better to deviate to R, whereafter player 2 finds it better to play r' after R, which in turn induces t_1 to choose L again. Continuing the argument two more steps we are back at the original equilibrium choices, hence, according to the critique no type of player 1 might have an incentive to deviate from L after all. This author's opinion is that the pooling outcome at L should not be considered self-enforcing: There are players that have an incentive to deviate. What the critique shows is that we do not know what will happen if it is suggested to the players to pool at L, but, as already seen above, equilibrium analysis cannot answer this question.

[Insert Figure 13 here]

In this author's opinion, the intuitive criteria that were discussed above may be criticized for the fact that they treat reached and unreached information sets asymmetrically: It is assumed that player 2 follows the recommendation after any message that is chosen

in equilibrium whereas he completely neglects the recommendation , and reoptimizes, after any unexpected message. To check self-enforcingness it is more appropriate to follow the symmetric procedure of first assuming that the recommendation is self-enforcing, that player 2 will always, i.e. after every message follow the recommendation, and then reject the recommendation if this assumption leads to a contradiction. Of course, this latter requirement is simply the INBR condition from the previous section. It is illustrated by means of the game of Figure 14. Cho and Kreps (1987) provide a similar example and claim that the elimination of the pooling equilibrium at L is not intuitive in this game.

[Insert Figure 14 here]

Consider the equilibrium outcome in which the types of player 1 pool at L. If we insist that recommendations be admissible (i.e. undominated) strategies, then to sustain pooling at L we should recommend that player 2 randomizes between m and r after R, putting at least half of the weight on r. Given this set of possible recommendations, choosing R is not a best response for type t_2, and after having eliminated this action, we see that player 2 prefers to choose l, hence, he wants to deviate from the recommendation. Consequently, if we insist on admissibility and INBR, then pooling at L cannot be self-enforcing. Note that none of the previous arguments discussed in this subsection, nor INBR alone, eliminates this outcome. (If the dominated strategy $\frac{2}{3}l + \frac{1}{3}r$ is allowed as a recommendation for player 2, then sending R is not inferior for type t_2.)

The literature also offers refined equilibrium notions that are not implied by stability. One such concept, that is frequently used in applications is that of *perfect sequential equilibrium* or PSE (Grossman and Perry (1986)). It is convenient to describe the slightly stronger notion of PSE* (Van Damme (1987)). Roughly, an equilibrium s fails to be a PSE* if there exists an unsent message m, a subset T' of types of player 1 and a response

r at m such that (i) if r is chosen at m then T' is exactly the set of types that prefer m to s and (ii) r is a best response against the conditional distribution of π on T'. (The formal definition is slightly different since types may be indifferent between deviating or not; such indifferences are handled as in (5.1), (5.2). The PSE concept is defined similarly but it is weaker since it allows player 1 to conjecture the 'wrong' response at m.) Hence, roughly, s fails to be a PSE* if there exists some message m and an equilibrium s' of the auxiliary game determined by s and m such that at least one type of player 1 prefers s' to s. Clearly, this concept is closely related to the forward induction requirement that was discussed at the end of the previous subsection. The difference is that there we required that there be a unique equilibrium that improves upon s, whereas here we allow there to be multiple improvements.

[Insert Figure 15 here]

Grossman and Perry (1986) have given an example to show that PSE need not exist. The game from Figure 15 shows that a stable set need not contain a PSE. In this game, pooling at L is stable but it is not a PSE. The outcome is stable since (roughly) stability allows player 2 to believe that any type might have deviated, hence, it allows player 2 to randomize in such a way that actually neither t_1 nor t_2 wants to deviate. The outcome is not a PSE since this concept forces player 2 to put weight 1 on either t_1 or t_2, hence, to choose either l or r. Clearly in either case at least one type of player 1 will want to deviate from L. This example makes clear that the PSE concept assumes that the players can coordinate their actions, i.e. that communication is possible and that communication indeed takes place. (However, note that player 2 has no incentive whatsoever to communicate.) Hence, PSE is not a purely noncooperative solution concept. In this author's opinion it is preferable to model communication explicitly by the rules of the game rather than indirectly by means of the solution concept. Such 'cheap talk' games

typically also have many equilibria and stability is not effective in reducing this set since every equilibrium outcome can be obtained by 'babbling', i.e. by using each message with positive probability. We will not consider cheap talk games any further; we just note that from the seminal papers Farrell (1985, 1990) and Grossman (1981) an extensive literature has sprung up, and that Matthews et al. (1990) survey the refinements used in this area. All these refinements assume that players will always accept the literal meaning of each statement unless it is logically contradictory, and the real challenge in this area seems to be to derive this assumption as a conclusion.

6 Equilibrium selection

Up to now we have dealt exclusively with the self-enforcingness aspect of equilibria, we did not discuss how self-enforcing norms come to be established nor how the selection among these takes place. We have seen, however, that considerations concerning self-enforcingness already lead to some conclusion concerning equilibrium selection: The basic idea of forward induction is that the equilibrium that is selected may depend on the context in which the game is played (cf. Figure 7.b). In this section we briefly discuss the approach to equilibrium selection and equilibrium attainment that is proposed in Carlsson and Van Damme (1990) (henceforth CD). CD picture players in the context in which the payoffs of the game are only "almost common knowledge" and they show that when a 2×2 game is played in this context, players reason themselves to the risk dominant equilibrium (Harsanyi and Selten, (1988)). (CD obtain results only for the class of 2×2 games. For general attacks on the equilibrium selection problem, see Harsanyi and Selten (1988) and Güth and Kalkofen (1989).)

[Insert Figure 16.a and 16.b here]

In the coordination game of Figure 16.a the equilibrium (L_1, L_2) satisfies the most strin-

gent requirements for self-enforcingness that have been discussed thus far: (L_1, L_2) is a strict equilibrium so that each player strictly looses by deviating if he expects the opponent to obey the recommendation to play this equilibrium. Of course, the question is whether a rational player will indeed expect his opponent to obey this recommendation. There is some evidence that at least human players do not consider such recommendations credible. Van Huyck et al. (1988) report on an experiment conducted with a 3×3 coordination game with (diagonal) payoffs (in dollarcents) of (90,90), (50,50), and (10,10) in which only 1 pair of players (out of 30) follows the recommendation to play (10,10): If (10,10) is recommended, then 47 of the 60 individuals (and 18 of the 30 pairs) deviate to the payoff dominant equilibrium (90,90). It is very likely that similar behavior would be observed in the game of Figure 15.a. One explanation for this behavior is that players are firmly convinced right from the start that only R makes sense in this game, that they consider any suggestion to play something else as being irrelevant and that such a suggestion can safely be ignored since it will be ignored by the opponent as well. The obvious question of course is how players can know that only R makes sense, and basically the answer that CD give is that players know this from reasoning through similar games. CD argue that the game from Figure 16.a should not be analyzed in isolation: Players know what to do in *this* game since they know that it is optimal to play the Pareto best equilibrium in *each* coordination game with Pareto ranked payoffs. CD suggest to analyze classes of games with the same structure simultaneously and they show that self-enforcing norms for how to play classes of games may prescribe a specific equilibrium of each element of the class, roughly because of the fact that norms will require that similar games be played similarly. (Fudenberg and Kreps (1988) present another approach to similarity in games. Of course the idea that a solution of a game should be part of a plan that is consistent across a larger domain occurs already in the seminal work of Nash (1950b) on bargaining and that of Schelling (1960) on focal points.)

The CD approach will now be illustrated by means of the game $\Gamma(\theta)$ from Figure 16.b. (The reader himself can supply the details for how the argument would run if the coordination game from Figure 16.a would be embedded in a one-dimensional parame-

ter of coordination games.) The game $\Gamma(7)$ has been extensively discussed in Aumann (1989). Aumann argues that if the players are convinced that they should play R then no amount of preplay communication can convince them to switch to L since each player knows that also a player who intends to play R will try to induce his opponent to switch to L. In $\Gamma(7)$ both L and R are strict equilibria and each one has something going for it: L Pareto dominates R but R is much safer. Hence, in this game there is a conflict between the intuitive notions of payoff dominance and risk dominance (Harsanyi and Selten (1988)). Formally, in a 2×2 game G, R is said to risk dominate L if the stability region of R (i.e. the set of all strategy vectors s against which R is a best reply) has a larger area than the stability region of L. Hence, in Figure 16.b, R risk dominates L if and only if $\theta > 4$. In their theory, Harsanyi and Selten resolve the conflict between the two intuitive notions in favor of payoff dominance; the reader should consult the postscript to their book for the arguments in favor of this choice.

Now imagine that the players are in the context in which they know that they have to play a game $\Gamma(\theta)$ as in Figure 16.b but they do not yet know which one. Hence, they know that they have to play a game in which the conflict between risk dominance and payoff dominance exists. (The reader may argue that the parametrization from Figure 16.b is not natural; We have chosen this parametrization to simplify the presentation. The assumptions to be discussed next are motivated similarly; the results from Carlsson and Van Damme (1990) are more general.) The reader will probably agree that as θ increases playing L becomes less and less attractive and that a natural way to play this game is by specifying a cutoff value $\hat{\theta}$ and play L if and only if θ is less than $\hat{\theta}$. CD show that, if the players can observe the actual parameter value θ only with some slight noise, then the value of $\hat{\theta}$ is uniquely determined in equilibrium. In fact $\hat{\theta} = 4$, hence, the players always choose the risk dominant equilibrium. (Note that some noise is essential to derive uniqueness, if θ could be perfectly observed, then each game $\Gamma(\theta)$ would occur as a simple subgame and the cutoff value may lie anywhere, in fact, in this case the equilibrium strategies need not be stepfunctions.)

To formally derive the above result let us assume that the set Θ of all possible parameter values is finite, that initially all values of θ are equally likely and that Θ includes values θ with $\theta < 0$ (which makes L_i strictly dominant) as well as values with $\theta > 8$ (such that R_i is strictly dominant). Furthermore, assume that, if the actual parameter value is θ, then one player receives the signal θ^+ (i.e. the smallest value in Θ that is larger than θ) while the other gets to hear θ^- (i.e. the largest value in Θ that is smaller than θ) with both possibilities being equally likely (with the appropriate modifications at the endpoint of Θ). Since the observations are noisy no player knows exactly which 'game' he is playing, however, if the grid of Θ is fine then each player has fairly accurate information about the payoffs in the game. Furthermore, in this case each player also has good knowledge about the information of his opponent and the players know that their perceptions of what the payoffs are, do not differ too much. Hence, if the grid of Θ is fine, the game with noisy observations may be viewed as a small perturbation of the game in which observations are perfect and in the latter $\Gamma(\theta)$ occurs as a subgame for each value of θ. However, it should be noted that from the point of view of common knowledge (Aumann (1976)), the games are completely different. Namely, in the unperturbed if a player receives the signal θ, then it is common knowledge that the game is $\Gamma(\theta)$, i.e. both players know that both players know ... that both players know that the game is $\Gamma(\theta)$. However, in the game with noise, if a player receives the signal θ, then he knows that the payoffs either are as in $\Gamma(\theta^-)$ or as in $\Gamma(\theta^+)$ and that his opponent either received the signal θ^{--} or θ^{++}. Hence, he also knows that the opponent believes that the game is either $\Gamma(\theta^{---})$ or $\Gamma(\theta^-)$ or $\Gamma(\theta^+)$ or $\Gamma(\theta^{+++})$, with all probabilities being equally likely, and that the opponent believes that his signal is either θ^{----}, or θ^{++++} or θ with the latter having probability $1/2$. Continuing inductively it is therefore seen that no matter how fine the grid size of Θ is, basically the only information that is common knowledge is that some game $\Gamma(\theta)$ with θ in Θ has to be played. This lack of common knowledge forces the players to take a global perspective in order to solve the perturbed game: To know what to do if one receives the signal θ one should also investigate what to do at parameter values θ' that are far away from θ. It is this phenomenon that drives the CD results. (A similar "action from a distance" also drives the results in Rubinstein's (1989)

electronic mail game.)

The analysis of the perturbed game is simple. Let θ_i be the observation of player i. If $\theta_i^+ < 0$ (resp. $\theta_i^- > 8$) then player i chooses L_i (resp. R_i) since he knows that this action is strictly dominant. Assume that it has already been shown by iterative elimination of strictly dominated strategies that L_1 and L_2 (resp. R_1 and R_2) are strictly dominant at each observation θ with $\theta \leq \alpha$ (resp. $\theta \geq \beta$). Hence, the iterative procedure starts with $\alpha = 0^{--}$ and $\beta = 8^{++}$. Consider $\theta_i = \alpha^+$, so that player i knows that either $\theta_j = \alpha^-$ or $\theta_j = \alpha^{+++}$, hence, player i knows that player j will choose L_j with a probability p that is at least $1/2$. Choosing L_i yields an expected payoff of $9p$ while R_i yields at most $\alpha^{++} + p$, so that player i will find it strictly dominant to choose L_i if $\alpha^{++} < 4$. Consequently, L_i is iteratively dominant for player i at θ_i if $\theta_i < 4^-$ and similarly R_i is iteratively dominant at θ_i if $\theta_i > 4^+$. We see that the perturbed game is almost dominance solvable: For all but a small set of parameter values (viz. the interval $[4^-, 4^+]$) unique iteratively dominant actions exist. By playing these dominant strategies players coordinate on the risk dominant equilibrium of the actual game that was selected by chance, hence, by just relying on rationalizability (Bernheim (1984), Pearce (1984)) in the perturbed game we obtain equilibrium selection according to the risk dominance criterion for every game $\Gamma(\theta)$ with $\theta \notin [4^-, 4^+]$.

Binmore (1990) has argued that in order to make progress in game theory it is necessary to model the way players think; that attention should be focused more on equilibriating processes rather than on equilibria. Although the model outlined above is rudimentary I believe that it captures some relevant aspects of reasoning processes. Certainly I do not want to claim the model's universal applicability; in some contexts the model may be relevant, in other contexts players may reason differently. The point, however, is that classical game theory is rich enough so as to provide models of the ways players might think.

References

Abdalla, A., R. Cooper, D. DeJong, R. Forsythe and T. Ross (1989). "Forward Induction in Coordination and Battle of the Sexes Games: Some Experimental Results", University of Iowa WP 89-22.

Aumann, R. (1974). "Subjectivity and correlation in randomized strategies", *Journal of Mathematical Economics*, **1**, 67-96.

Aumann, R. (1976). "Agreeing to Disagree", *Annals of Statistics*, **4**, 1236-1239.

Aumann, R. (1987a). "Correlated Equilibrium as an Expression of Bayesian Rationality", *Econometrica*, **55**, 1-18.

Aumann, R. (1987b). "What is Game Theory Trying to Accomplish?", in *Frontiers of Economics*, eds. K. Arrow and S. Honkapohja, 28-100.

Aumann, R. (1988). "Irrationality in Game Theory", Paper for conference on Economic Theories of Politics.

Aumann, R. (1989). "Nash Equilibria are Not Self-Enforcing", Mimeo, Hebrew University of Jerusalem.

Bagwell, K. and G. Ramey (1990). "Capacity, Entry and Forward Induction", University of California at San Diego DP 90-22.

Banks, J.S. and J. Sobel (1987). "Equilibrium Selection in Signalling Games", *Econometrica*, **55**, 647-661.

Basu, K. (1988). "Strategic Irrationality in Extensive Games", *Mathematical Social Sciences*, **15**, 247-260.

Basu, K. (1990). "On the Non-Existence of Rationality Definition for Extensive Games", *International Journal of Game Theory*, **19**, 33-44.

Ben-Porath, E. and E. Dekel (1987). "Coordination and the Potential for Self-Sacrifice", Research Paper No. 984, Graduate School of Business, Stanford University.

Benoit, J.-P. and V. Krishna (1985). "Finitely Repeated Games", *Econometrica*, **53**, 905-922.

Bernheim, D. (1984). "Rationalizable Strategic Behavior", *Econometrica*, **52**, 1007-1028.

Bernheim, D. (1986). "Axiomatic Characterizations of Rational Choice in Strategic Environments", *Scandinavian Journal of Economics*, **88**, 473-488.

Binmore, K. (1987). "Modeling Rational Players I", *Economics and Philosophy*, **3**, 179-214.

Binmore, K. (1988). "Modeling Rational Players, II", *Economics and Philosophy*, **4**, 9-55.

Binmore, K. (1990). "Foundations of Game Theory", Lecture delivered at the 6th World Congress of the Econometric Society in Barcelona.

Brandenburger, A. and E. Dekel (1987). "Rationalizability and Correlated Equilibria", *Econometrica*, **55**, 1391-1402.

Canning, D. (1989). "Convergence to Equilibrium in a Sequence of Games with Learning", STICERD Discussion Paper.

Canning, D. (1990). "Social Equilibrium", Mimeo, Pembroke College, Cambridge.

Carlsson, H. and E. van Damme (1990). "Global Games and Equilibrium Selection", CentER DP 9052, Tilburg University.

Cho, I.K. (1987). "A Refinement of Sequential Equilibrium", *Econometrica*, **55**, 1367-1390.

Cho, I.K. (1990). "Strategic Stability in Repeated Signaling Games", Mimeo, University of Chicago.

Cho, I.K. and D.M. Kreps (1987). "Signalling Games and Stable Equilibria", *Quarterly Journal of Economics*, **102**, 179-221.

Cho, I.K. and J. Sobel (1990). "Strategic Stability and Uniqueness in Signaling Games", *Journal of Economic Theory*, **50**, 381-413.

Dekel, E. (1988). "Simultaneous Offers and the Inefficiency of Bargaining: A Two-period Example", Mimeo, University of California at Berkeley, forthcoming in *Journal of Economic Theory*.

Farrell, J. (1984). "Credible Neologisms in Games and Communication", Mimeo, Massachusetts Institute of Technology.

Farrell, J. (1990). "Meaning and Credibility in Cheap-Talk Games", forthcoming in *Mathematical Models in Economics*, ed. M. Dempster, Oxford University Press.

Forges, F. (1986). "An Approach to Communication Equilibria", *Econometrica*, **54**, 1375-1385.

Fudenberg, D. and D. Kreps (1988). "A Theory of Learning Experimentation and Equilibrium in Games", Manuscript.

Fudenberg, D., D. Kreps and D. Levine (1988). "On the Robustness of Equilibrium Refinements", *Journal of Economic Theory*, **44**, 354-380.

Fudenberg, D. and E. Maskin (1986). "The Folk Theorem in Repeated Games with Discounting and with Incomplete Information", *Econometrica*, **54**, 533-554.

Fudenberg, D. and J. Tirole (1989). "Perfect Bayesian Equilibrium and Sequential Equilibrium", Mimeo, Harvard University, forthcoming in *Journal of Economic Theory*.

Glazer, J. and A. Weiss (1990). "Pricing and Coordination: Strategically Stable Equilibrium", *Games and Economic Behavior*, **2**, 118-128.

Grossman, S. (1981). "The Informational Role of Warranties and the Private Disclosure about Product Quality", *Journal of Law and Economics*, **24**, 461-483.

Grossman, S. and M. Perry (1986). "Perfect Sequential Equilibrium", *Journal of Economic Theory*, **39**, 97-119.

Güth, W. and B. Kalkofen (1989). "Unique Solutions for Strategic Games", *Lecture Notes in Economics and Mathematical Systems*, **328**, Springer Verlag, Berlin.

Harsanyi, J. (1973). "Games with Randomly Disturbed Payoffs: A New Rationale for Mixed Strategy Equilibrium Points", *International Journal of Game Theory*, **2**, 1-23.

Harsanyi, J. and R. Selten (1988). *A General Theory of Equilibrium Selection in Games*, MIT Press, Cambridge, MA.

Hillas, J. (1990). "On the Definition of the Strategic Stability of Equilibria", *Econometrica*, **58**, 1365-1390.

Kalai, E. and E. Lehrer (1990). "Rational Learning Leads to Nash Equilibrium", Mimeo, Northwestern University.

Kalai, E. and D. Samet (1984). "Persistent Equilibria", *International Journal of Game Theory*, **13**, 129-141.

Kohlberg, E. (1989). "Refinement of Nash Equilibrium: The Main Ideas", Mimeo, Harvard University.

Kohlberg, E. and J.-F. Mertens (1986). "On the Strategic Stability of Equilibria", *Econometrica*, **54**, 1003-1037.

Kohlberg, E. and P. Reny (1991). "Consistent Assessments in Sequential Equilibria", unpublished notes.

Kreps, D. and G. Ramey (1987). "Structural Consistency, Consistency, and Sequential Rationality", *Econometrica*, **55**, 1331-1348.

Kreps, D., P. Milgrom, J. Roberts and R. Wilson (1982). "Rational Cooperation in the Finitely-Repeated Prisoners' Dilemma", *Journal of Economic Theory*, **27**, 245-252.

Kreps, D. and R. Wilson (1982a). "Sequential Equilibria", *Econometrica*, **50**, 863-894.

Kreps, D. and R. Wilson (1982b). "Reputation and Imperfect Information", *Journal of Economic Theory*, **27**, 253-279.

Kreps, D. and J. Sobel (forthcoming). "Signalling", in *Handbook of Game Theory* eds. R. Aumann and S. Hart.

Lewis, D. (1973). *Counterfactuals*, Blackwell, Oxford.

Luce, R. and H. Raiffa (1957). *Games and Decisions*, Wiley, New York.

Madrigal, V., T. Tan and S. Werlang (1987). "Support Restrictions and Sequential Equilibria", *Journal of Economic Theory*, **43**, 329-334.

Maskin, E. and J. Tirole (1989). "Markov Equilibrium", mimeo, Harvard University.

Matthews, S., M. Okuno-Fujiwara and A. Postlewaite (1990). "Refining Cheap-Talk Equilibria", Mimeo, Northwestern University.

Maynard Smith, J. (1982). *Evolution and the Theory of Games*, Cambridge University Press.

Maynard Smith, J. and G. Price (1973). "The Logic of Animal Conflict", *Nature*, London, **246**, 15-18.

McLennan, A. (1985). "Justifiable Beliefs in Sequential Equilibrium", *Econometrica*, **53**, 889-904.

Mertens, J.-F. (1987). "Ordinality in Non Cooperative Games", CORE Discussion Paper No. 8728.

Mertens, J.-F. (1988). "Stable Equilibria – A Reformulation", CORE Discussion Paper No. 8838.

Mertens, J.-F. (1989a). "Stable Equilibria - A Reformulation, Part I", *Mathematics of Operations Research*, **14**, 575-625.

Mertens, J.-F. (1989b). "Equilibrium and Rationality: Context and History-Dependence", Mimeo, CORE.

Mertens, J.-F. (1990). "The 'Small Worlds' Axiom for Stable Equilibria", CORE DP 9007.

Milgrom, P. and J. Roberts (1989). "Adaptive and Sophisticated Learning in Repeated Normal Form Games", mimeo Stanford University. To appear in *Games and Economic Behavior.*

Milgrom, P. and J. Roberts (1990). "Rationalizability, Learning, and Equilibrium in Games with Strategic Complementarities", *Econometrica*, **58**, 1255-1277.

Myerson, R. (1978). "Refinements of the Nash Equilibrium Concept", *International Journal of Game Theory*, **7**, 73-80.

Myerson, R. (1986). "Multistage Games with Communication", *Econometrica*, **54**, 323-358.

Myerson, R. (1989). "Credible Negotiation Statements and Coherent Plans", *Journal of Economic Theory*, **48**, 264-303.

Nash, J. (1950a). "Equilibrium Points in n-person Games", *Proceedings from the National Academy of Sciences, U.S.A.*, **36**, 48-49.

Nash, J. (1950b). "The Bargaining Problem", *Econometrica*, **18**, 155-162.

Nöldeke, G. and E. van Damme (1990). "Switching Away From Probability One Beliefs", University of Bonn DP A-304.

Osborne, M. (1990). "Signaling, Forward Induction, and Stability in Finitely Repeated Games", *Journal of Economic Theory*, **50**,, 22-36.

Pearce, D. (1984). "Rationalizable Strategic Behavior and the Problem of Perfection", *Econometrica*, **52**, 1029-1050.

Pearce, D. (1990). "Renegotiation in Repeated Games", Lecture delivered at the 6th World Congress of the Econometric Society in Barcelona.

Ponssard, J.-P. (1990a). "Self Enforceable Paths in Games in Extensive Form: A Behavioral Approach Based on Interactivity", *Theory and Decision*, **28**, 69-83.

Ponssard, J.-P. (1990b). "Forward Induction and Sunk Costs Give Average Cost Pricing", forthcoming in *Games and Economic Behavior*.

Ponssard, J.-P. [1990c]. "A Note on Forward Induction and Escalation Games with Perfect Information", Mimeo, Ecole Polytechnique, Paris.

Reny, P. (1988a). "Backward Induction, Normal Form Perfection and Explicable Equilibria", Mimeo, University of Western Ontario.

Reny, P. (1988b). "Backward Induction and Common Knowledge in Games with Perfect Information", Mimeo, University of Western Ontario.

Rosenthal, R. (1981). "Games of Perfect Information, Predatory Pricing, and the Chain-Store Paradox", *Journal of Economic Theory*, **25**, 92-100.

Rubinstein, A. (1988). "Comments on the Interpretation of Game Theory", STICERD Discussion Paper. Forthcoming in *Econometrica*.

Rubinstein, A. (1989). "The Electronic Mail Game: Strategic Behavior Under 'Almost Common Knowledge' ", *American Economic Review*, **79**,

385-391.

Schelling, T.C. (1960). *The Strategy of Conflict*, Harvard University Press, Cambridge, MA.

Selten, R. (1965). "Spieltheoretische Behandlung eines Oligopolmodels mit Nachfragetragheit", *Zeitschrift für die Gesamte Staatswissenschaft*, **12**, 301-324 and 667-689.

Selten, R. (1975). "Re-examination of the Perfectness Concept for Equilibrium Points in Extensive Games', *International Journal of Game Theory*, **4**, 25-55.

Selten, R. and U. Leopold (1982). "Subjunctive Conditionals in Decision Theory and Game Theory", in *Studies in Economics* eds. Stegmuller, Balzer and Spohn, Springer Verlag, Berlin. *Philosophy of Economics*, Vol. 2.

Sobel, J., L. Stole and I. Zapater (1990). "Fixed-Equilibrium Rationalizability in Signalling Games", Discussion Paper 90-13, University of California, San Diego.

Spence, M. (1974). *Market Signalling*, Harvard University Press, Cambridge MA.

Stalnaker, R. (1969). "A Theory of Conditionals", in *Studies in Logical Theory* ed. N. Rescher, Blackwell, Oxford.

Suehiro, H. (1990). "On a 'mistaken theories' refinement", Mimeo, Kobe University.

Tan, T. and S. Werlang (1988). "The Bayesian Foundations of Solution Concepts of Games", *Journal of Economic Theory*, **45**, 370-391.

Van Damme, E. (1984). "A Relation Between Perfect Equilibria in Extensive Form Games and Proper Equilibria in Normal Form Games", *International Journal of Game Theory*, **13**, 1-13.

Van Damme, E. (1987). *Stability and Perfection of Nash Equilibria*, Springer-Verlag, Berlin.

Van Damme, E. (1988). "The Impossibility of Stable Renegotiation", *Economic Letters*, **26**, 321-324.

Van Damme, E. (1989). "Stable Equilibria and Forward Induction", *Journal of Economic Theory*, **48**, 476-496.

Van Huyck, J.B., A.B. Gillette and R.C. Battalio (1988). "Credible Assignments in Non-Cooperative Games", Texas A&M University working paper.

Von Neumann, J, and O. Morgenstern (1948). *Theory of Games and Economic Behavior*, Princeton University Press, Princeton NJ.

Vincent, D. (1990). "Bilateral Monopoly, Non-durable Goods and Dynamic Trading Relationships", Mimeo, Northwestern University.

Weibull, J. (1990). "On Self-Enforcement in Extensive-Form Games", Mimeo, Princeton University.

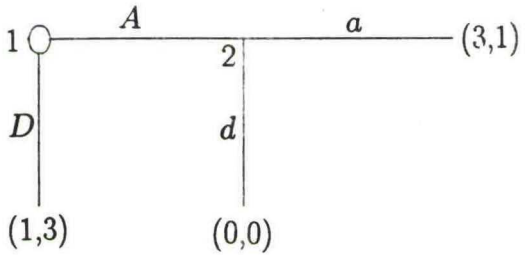


Figure 1.a: game Γ_1.

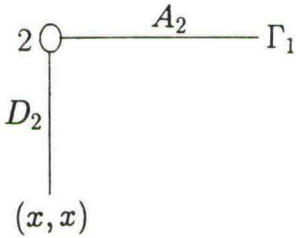


Figure 1.b: game $\Gamma_2(x)$.

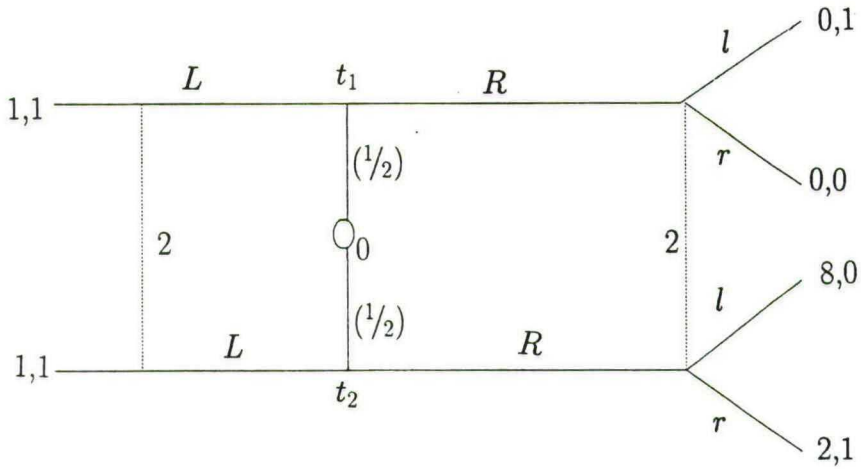
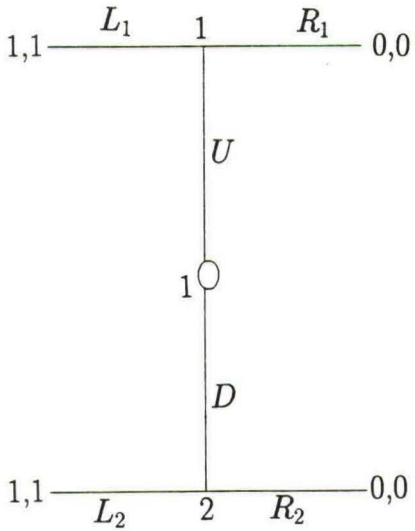


Figure 2.a: A signalling game

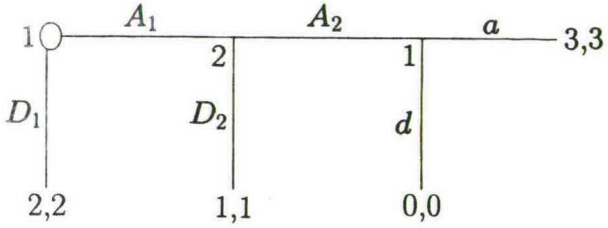
			l	r
1,1		$(t_1, 1/2)$	0,1	0,0
1,1		$(t_2, 1/2)$	8,0	2,1
L			R	

Figure 2.b: A matrix representation of the signalling game from Figure 2.a.



	L_2	R_2
UL_1	1,1	1,1
UR_1	0,0	0,0
D	1,1	0,0

Figure 3: An extensive form perfect equilibrium need not be admissible.



	D_2	A_2
D_1	2,2	2,2
A_1d	1,1	0,0
A_1a	1,1	3,3

Figure 4: Normal form perfectness does not imply subgame perfection.

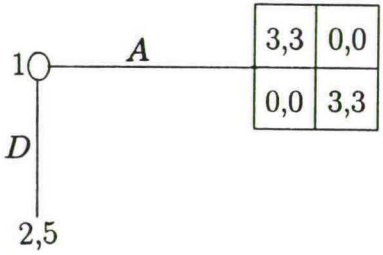


Figure 5: A coordination problem

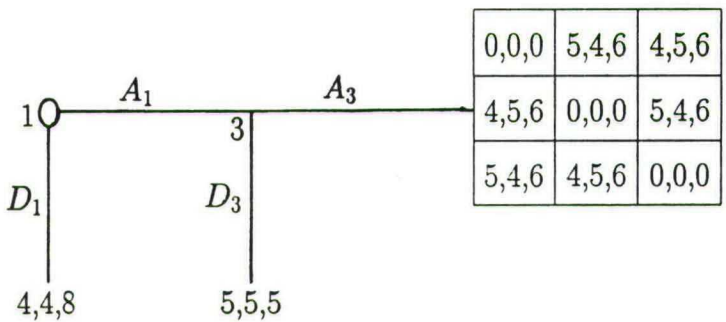


Figure 6: Correlation may yield non-subgame perfect equilibrium outcomes.

	s	w
s	0,0	3,1
w	1,3	0,0

Figure 7.a: Battle of the Sexes (BS)

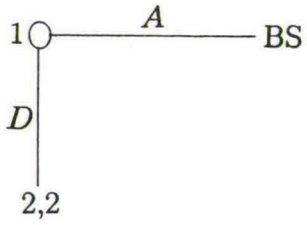


Figure 7.b: BS with an outside option.

	L	R
l	6,6	6,6
m	2,0	0,2
r	0,2	2,0

Figure 8.a.

	L	R
l	6,6	6,6
s	4,3	3,4
m	2,0	0,2
r	0,2	2,0

Figure 8.b.

Figure 8: Extensive form games with the same reduced normal form with disjoint sets of SPE.

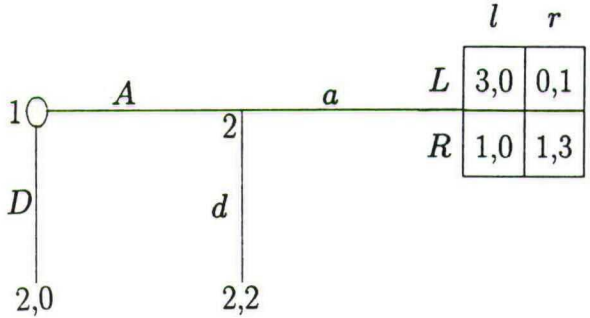


Figure 9: Different elimination orders yield different outcomes.

			l	r
2,2		$(t_1, 1/2)$	3,3	0,0
2,2		$(t_2, 1/2)$	0,0	1,1
L			R	

Figure 10: Pooling at L is not 'intuitive'.

		l	m	r
2,2	$(t_1, 1/3)$	5,5	0,0	0,0
2,2	$(t_2, 1/3)$	0,0	5,5	0,0
2,2	$(t_3, 1/3)$	0,0	0,0	1,1
L		R		

Figure 11: Equilibrium dominance is more restrictive than the intuitive criterion.

		l	r	l'	r'
2,2	$(t_1, 1/2)$	4,4	1,1	5,0	0,1
2,2	$(t_2, 1/2)$	5,0	0,1	4,4	1,1
L		M		R	

Figure 12: Investigating unsent messages separately or simultaneously may make a difference.

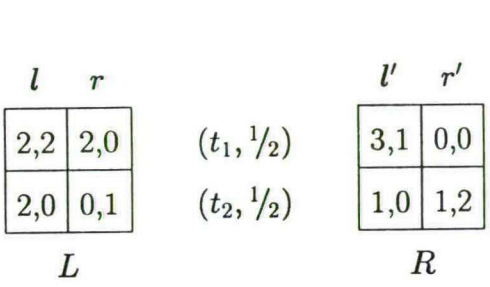


Figure 13: Illustration of the 'Stiglitz critique'.

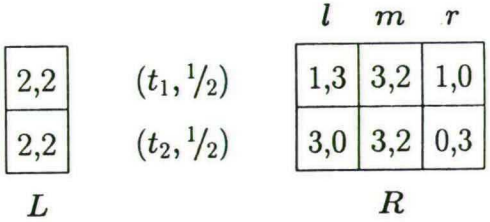


Figure 14: The combination of admissibility and INBR eliminates pooling at L.

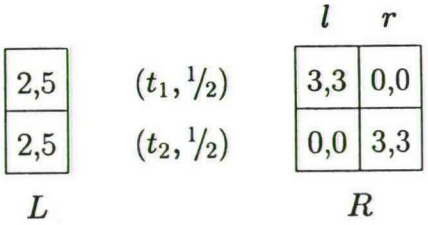


Figure 15: Pooling at L is stable but is not a PSE.

	L_2	R_2
L_1	1,1	0,0
R_1	0,0	2,2

Figure 16.a: A coordination Game.

	L_2	R_2
L_1	9,9	$0, 1+\theta$
R_1	$1+\theta, 0$	θ, θ

Figure 16.b: Game $\Gamma(\theta)$.

Discussion Paper Series, CentER, Tilburg University, The Netherlands:

(For previous papers please consult previous discussion papers.)

No.	Author(s)	Title
8954	A. Kapteyn, S. van de Geer, H. van de Stadt and T. Wansbeek	Interdependent Preferences: An Econometric Analysis
8955	L. Zou	Ownership Structure and Efficiency: An Incentive Mechanism Approach
8956	P.Kooreman and A. Kapteyn	On the Empirical Implementation of Some Game Theoretic Models of Household Labor Supply
8957	E. van Damme	Signaling and Forward Induction in a Market Entry Context
9001	A. van Soest, P. Kooreman and A. Kapteyn	Coherency and Regularity of Demand Systems with Equality and Inequality Constraints
9002	J.R. Magnus and B. Pesaran	Forecasting, Misspecification and Unit Roots: The Case of AR(1) Versus ARMA(1,1)
9003	J. Driffill and C. Schultz	Wage Setting and Stabilization Policy in a Game with Renegotiation
9004	M. McAleer, M.H. Pesaran and A. Bera	Alternative Approaches to Testing Non-Nested Models with Autocorrelated Disturbances: An Application to Models of U.S. Unemployment
9005	Th. ten Raa and M.F.J. Steel	A Stochastic Analysis of an Input-Output Model: Comment
9006	M. McAleer and C.R. McKenzie	Keynesian and New Classical Models of Unemployment Revisited
9007	J. Osiewalski and M.F.J. Steel	Semi-Conjugate Prior Densities in Multi-variate t Regression Models
9008	G.W. Imbens	Duration Models with Time-Varying Coefficients
9009	G.W. Imbens	An Efficient Method of Moments Estimator for Discrete Choice Models with Choice-Based Sampling
9010	P. Deschamps	Expectations and Intertemporal Separability in an Empirical Model of Consumption and Investment under Uncertainty
9011	W. Güth and E. van Damme	Gorby Games - A Game Theoretic Analysis of Disarmament Campaigns and the Defense Efficiency-Hypothesis
9012	A. Horsley and A. Wrobel	The Existence of an Equilibrium Density for Marginal Cost Prices, and the Solution to the Shifting-Peak Problem

No.	Author(s)	Title
9013	A. Horsley and A. Wrobel	The Closedness of the Free-Disposal Hull of a Production Set
9014	A. Horsley and A. Wrobel	The Continuity of the Equilibrium Price Density: The Case of Symmetric Joint Costs, and a Solution to the Shifting-Pattern Problem
9015	A. van den Elzen, G. van der Laan and D. Talman	An Adjustment Process for an Exchange Economy with Linear Production Technologies
9016	P. Deschamps	On Fractional Demand Systems and Budget Share Positivity
9017	B.J. Christensen and N.M. Kiefer	The Exact Likelihood Function for an Empirical Job Search Model
9018	M. Verbeek and Th. Nijman	Testing for Selectivity Bias in Panel Data Models
9019	J.R. Magnus and B. Pesaran	Evaluation of Moments of Ratios of Quadratic Forms in Normal Variables and Related Statistics
9020	A. Robson	Status, the Distribution of Wealth, Social and Private Attitudes to Risk
9021	J.R. Magnus and B. Pesaran	Evaluation of Moments of Quadratic Forms in Normal Variables
9022	K. Kamiya and D. Talman	Linear Stationary Point Problems
9023	W. Emons	Good Times, Bad Times, and Vertical Upstream Integration
9024	C. Dang	The D_2-Triangulation for Simplicial Homotopy Algorithms for Computing Solutions of Nonlinear Equations
9025	K. Kamiya and D. Talman	Variable Dimension Simplicial Algorithm for Balanced Games
9026	P. Skott	Efficiency Wages, Mark-Up Pricing and Effective Demand
9027	C. Dang and D. Talman	The D_1-Triangulation in Simplicial Variable Dimension Algorithms for Computing Solutions of Nonlinear Equations
9028	J. Bai, A.J. Jakeman and M. McAleer	Discrimination Between Nested Two- and Three-Parameter Distributions: An Application to Models of Air Pollution
9029	Th. van de Klundert	Crowding out and the Wealth of Nations

No.	Author(s)	Title
9030	Th. van de Klundert and R. Gradus	Optimal Government Debt under Distortionary Taxation
9031	A. Weber	The Credibility of Monetary Target Announcements: An Empirical Evaluation
9032	J. Osiewalski and M. Steel	Robust Bayesian Inference in Elliptical Regression Models
9033	C. R. Wichers	The Linear-Algebraic Structure of Least Squares
9034	C. de Vries	On the Relation between GARCH and Stable Processes
9035	M.R. Baye, D.W. Jansen and Q. Li	Aggregation and the "Random Objective" Justification for Disturbances in Complete Demand Systems
9036	J. Driffill	The Term Structure of Interest Rates: Structural Stability and Macroeconomic Policy Changes in the UK
9037	F. van der Ploeg	Budgetary Aspects of Economic and Monetary Integration in Europe
9038	A. Robson	Existence of Nash Equilibrium in Mixed Strategies for Games where Payoffs Need not Be Continuous in Pure Strategies
9039	A. Robson	An "Informationally Robust Equilibrium" for Two-Person Nonzero-Sum Games
9040	M.R. Baye, G. Tian and J. Zhou	The Existence of Pure-Strategy Nash Equilibrium in Games with Payoffs that are not Quasiconcave
9041	M. Burnovsky and I. Zang	"Costless" Indirect Regulation of Monopolies with Substantial Entry Cost
9042	P.J. Deschamps	Joint Tests for Regularity and Autocorrelation in Allocation Systems
9043	S. Chib, J. Osiewalski and M. Steel	Posterior Inference on the Degrees of Freedom Parameter in Multivariate-t Regression Models
9044	H.A. Keuzenkamp	The Probability Approach in Economic Methodology: On the Relation between Haavelmo's Legacy and the Methodology of Economics
9045	I.M. Bomze and E.E.C. van Damme	A Dynamical Characterization of Evolutionarily Stable States
9046	E. van Damme	On Dominance Solvable Games and Equilibrium Selection Theories

No.	Author(s)	Title
9047	J. Driffill	Changes in Regime and the Term Structure: A Note
9048	A.J.J. Talman	General Equilibrium Programming
9049	H.A. Keuzenkamp and F. van der Ploeg	Saving, Investment, Government Finance and the Current Account: The Dutch Experience
9050	C. Dang and A.J.J. Talman	The D_1-Triangulation in Simplicial Variable Dimension Algorithms on the Unit Simplex for Computing Fixed Points
9051	M. Baye, D. Kovenock and C. de Vries	The All-Pay Auction with Complete Information
9052	H. Carlsson and E. van Damme	Global Games and Equilibrium Selection
9053	M. Baye and D. Kovenock	How to Sell a Pickup Truck: "Beat-or-Pay" Advertisements as Facilitating Devices
9054	Th. van de Klundert	The Ultimate Consequences of the New Growth Theory; An Introduction to the Views of M. Fitzgerald Scott
9055	P. Kooreman	Nonparametric Bounds on the Regression Coefficients when an Explanatory Variable is Categorized
9056	R. Bartels and D.G. Fiebig	Integrating Direct Metering and Conditional Demand Analysis for Estimating End-Use Loads
9057	M.R. Veall and K.F. Zimmermann	Evaluating Pseudo-R^2's for Binary Probit Models
9058	R. Bartels and D.G. Fiebig	More on the Grouped Heteroskedasticity Model
9059	F. van der Ploeg	Channels of International Policy Transmission
9060	H. Bester	The Role of Collateral in a Model of Debt Renegotiation
9061	F. van der Ploeg	Macroeconomic Policy Coordination during the Various Phases of Economic and Monetary Integration in Europe
9062	E. Bennett and E. van Damme	Demand Commitment Bargaining: - The Case of Apex Games
9063	S. Chib, J. Osiewalski and M. Steel	Regression Models under Competing Covariance Matrices: A Bayesian Perspective
9064	M. Verbeek and Th. Nijman	Can Cohort Data Be Treated as Genuine Panel Data?

No.	Author(s)	Title
9065	F. van der Ploeg and A. de Zeeuw	International Aspects of Pollution Control
9066	F.C. Drost and Th. E. Nijman	Temporal Aggregation of GARCH Processes
9067	Y. Dai and D. Talman	Linear Stationary Point Problems on Unbounded Polyhedra
9068	Th. Nijman and R. Beetsma	Empirical Tests of a Simple Pricing Model for Sugar Futures
9069	F. van der Ploeg	Short-Sighted Politicians and Erosion of Government Assets
9070	E. van Damme	Fair Division under Asymmetric Information
9071	J. Eichberger, H. Haller and F. Milne	Naive Bayesian Learning in 2 x 2 Matrix Games
9072	G. Alogoskoufis and F. van der Ploeg	Endogenous Growth and Overlapping Generations
9073	K.C. Fung	Strategic Industrial Policy for Cournot and Bertrand Oligopoly: Management-Labor Cooperation as a Possible Solution to the Market Structure Dilemma
9101	A. van Soest	Minimum Wages, Earnings and Employment
9102	A. Barten and M. McAleer	Comparing the Empirical Performance of Alternative Demand Systems
9103	A. Weber	EMS Credibility
9104	G. Alogoskoufis and F. van der Ploeg	Debts, Deficits and Growth in Interdependent Economies
9105	R.M.W.J. Beetsma	Bands and Statistical Properties of EMS Exchange Rates
9106	C.N. Teulings	The Diverging Effects of the Business Cycle on the Expected Duration of Job Search
9107	E. van Damme	Refinements of Nash Equilibrium